

Distinct timescales of population coding across cortex

Caroline A. Runyan^{1*}, Eugenio Piasini^{2*}, Stefano Panzeri² & Christopher D. Harvey¹

The cortex represents information across widely varying timescales^{1–5}. For instance, sensory cortex encodes stimuli that fluctuate over few tens of milliseconds^{6,7}, whereas in association cortex behavioural choices can require the maintenance of information over seconds^{8,9}. However, it remains poorly understood whether diverse timescales result mostly from features intrinsic to individual neurons or from neuronal population activity. This question remains unanswered, because the timescales of coding in populations of neurons have not been studied extensively, and population codes have not been compared systematically across cortical regions. Here we show that population codes can be essential to achieve long coding timescales. Furthermore, we find that the properties of population codes differ between sensory and association cortices. We compared coding for sensory stimuli and behavioural choices in auditory cortex and posterior parietal cortex as mice performed a sound localization task. Auditory stimulus information was stronger in auditory cortex than in posterior parietal cortex, and both regions contained choice information. Although auditory cortex and posterior parietal cortex coded information by tiling in time neurons that were transiently informative for approximately 200 milliseconds, the areas had major differences in functional coupling between neurons, measured as activity correlations that could not be explained by task events. Coupling among posterior parietal cortex neurons was strong and extended over long time lags, whereas coupling among auditory cortex neurons was weak and short-lived. Stronger coupling in posterior parietal cortex led to a population code with long timescales and a representation of choice that remained consistent for approximately 1 second. In contrast, auditory cortex had a code with rapid fluctuations in stimulus and choice information over hundreds of milliseconds. Our results reveal that population codes differ across cortex and that coupling is a variable property of cortical populations that affects the timescale of information coding and the accuracy of behaviour.

The goal of this work was to compare coding across cortical regions for two key features of behavioural tasks: stimulus and choice. We developed a sound localization task in which mice reported perceptual decisions by navigating through a visual virtual reality T-maze¹⁰ (Fig. 1a). As mice ran down the T-stem, a sound cue was played from one of eight possible locations in head-centred, real-world coordinates. Mice reported whether the sound originated from their left or right by turning in that direction at the T-intersection (Fig. 1b, c).

We focused on auditory cortex (AC), because it is necessary for sound localization tasks¹¹ and on posterior parietal cortex (PPC), because it is involved in spatial auditory processing¹², receives inputs from AC¹⁰, is a multisensory-motor interface^{8–10,13–17}, and is essential for virtual-navigation tasks¹⁰. In each mouse, we imaged the activity of ~50 neurons simultaneously from AC and PPC on separate days. In both regions, neurons were transiently active at different time points, resulting in a population that tiled the trial (Fig. 1d, e).

Activity in some AC neurons was selective for stimulus location; however, as a population, AC activity was heterogeneous and complex (Fig. 1d, f, Extended Data Fig. 1 and Supplementary Information). In PPC, a substantial fraction of neurons had different activity on trials with opposite behavioural choices^{8–10,15–18} (Fig. 1f).

The heterogeneity of activity patterns suggested that, in addition to stimulus and choice, multiple task-related variables, such as visual inputs, reward delivery, and the mouse's running, might affect neuronal responses. To take these variables into account and to help isolate signals related to stimulus and choice, we developed an encoding model (generalized linear model, GLM). This model incorporated all measured task-related variables as predictors of each neuron's activity^{19,20} (Fig. 2a and Extended Data Fig. 2a–d). The model reliably predicted the time course and selectivity of single-neuron activity in AC and PPC (Fig. 2b, c and Extended Data Fig. 2g).

To determine whether stimulus and choice information were present in AC and PPC, we decoded the most likely stimulus category (left or right location) or choice (left or right turn) from neuronal activity by using Bayes' rule to invert the prediction of the encoding model (Fig. 2a).

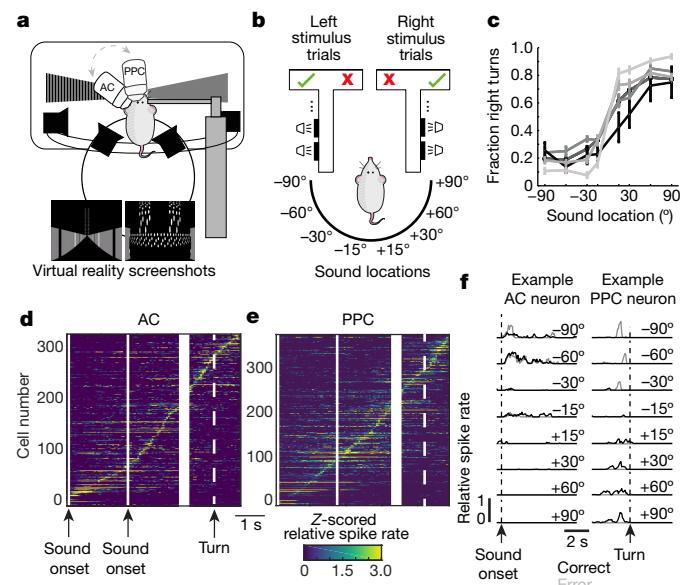


Figure 1 | Imaging AC and PPC responses during a sound localization task. **a**, Schematic of experimental set-up with virtual reality screenshots for the beginning of the trial and T-intersection. **b**, Schematic of the task. Left/right sound category (speaker symbols) indicate the rewarded side of the maze (checkmark). **c**, Behavioural performance. Each line is the average across sessions for a single mouse; error bars, s.e.m.; $n = 5$ mice. **d–e**, Z-scored, trial-averaged activity of all AC ($n = 329$) and PPC neurons ($n = 386$), sorted by time of peak mean activity. **f**, Average responses of example AC and PPC neurons on correct (black) and error (grey) trials.

¹Department of Neurobiology, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Neural Computation Laboratory, Istituto Italiano di Tecnologia, 38068 Rovereto, Italy.

*These authors contributed equally to this work.

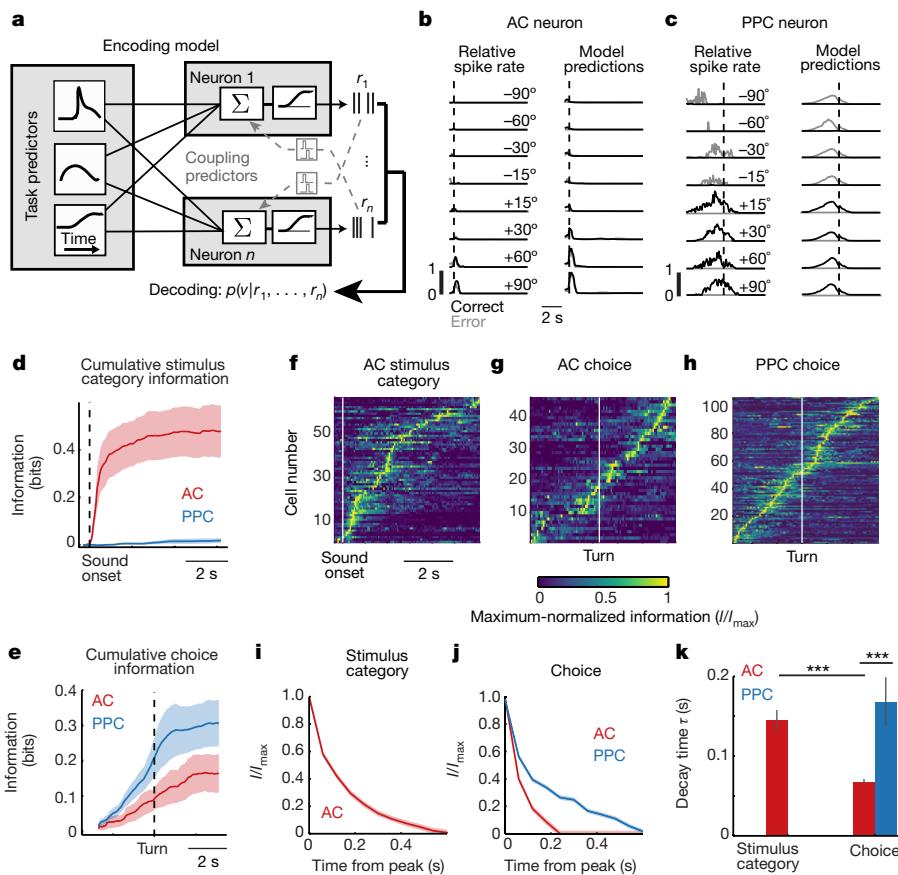


Figure 2 | Encoding and decoding stimulus and choice information in AC and PPC. **a**, The encoding model (GLM) was fitted to each neuron's activity using task predictors ('uncoupled model') or task predictors and activity of other neurons (coupling predictors, 'coupled model'). To decode stimuli or choices (indicated as v), the posterior probability of each stimulus or choice was computed, on the basis of the current time point only ('instantaneous decoder'), or all previous time points in the trial ('cumulative decoder'). The decoder could use either all neurons (population decoder) or individual neurons (single-neuron decoder). **b, c**, Example trial-averaged responses (left column) and model predictions (right column) in correct (black) and error (grey) trials. **d, e**, Cumulative stimulus category and choice information decoded in AC (red) and PPC (blue) populations. Data are mean \pm s.e.m. across

Because stimulus locations and choices were related to one another by task design (for example, left stimuli required a left turn for reward), we decoupled these features and isolated information purely related to stimulus from information purely related to choice by selecting equal numbers of right and left choice trials for analysis in each stimulus condition (thus the same number of correct and error trials). Decoding performance was calculated as the mutual information between the decoded and actual stimulus category or choice.

Pure information about the stimulus category was present in AC activity but was weak in PPC (Fig. 2d). Information purely about the choice was present in both AC and PPC populations (Fig. 2e). Although the mouse's running patterns that triggered turns in the virtual environment necessarily covaried with the choice at some trial time points, the choice information estimated by our decoding analysis was similar even when fully discounting the effects of running patterns, suggesting that AC and PPC contained genuine choice information (Extended Data Fig. 3g–i). Additional analyses that did not decouple stimulus and choice revealed that both regions had a relationship between neuronal tuning for stimulus and choice and contained information at the intersection between stimulus and choice²¹ (Supplementary Information, Extended Data Figs 3a–f and 4a). The information present in both regions thus appeared to be used for performing the task.

datasets; $n = 7$ datasets for AC and PPC. **f–h**, Maximum-normalized instantaneous information about stimulus category and choice for AC and PPC cells with at least 0.06 bits of information at any point in the trial (fraction of all cells: AC: stimulus, $20.2 \pm 5.4\%$; choice, $15.9 \pm 3.2\%$; PPC: stimulus, $3.1 \pm 1.3\%$; choice, $19.3 \pm 4.6\%$). Neurons were sorted by the times of their maximum information. **i, j**, Information about stimulus category and choice averaged across all AC (red) and PPC (blue) cells, calculated as a two-sided decay (forwards and backwards in time) around the information peak of each cell. Data are mean \pm s.e.m. across cells with at least 0.06 bits of information. **k**, Decay time constant of single-cell information from exponential fits to information time courses in **i** and **j**. Error bars, 95% confidence intervals (Methods). *** $P < 0.001$, z-test.

We investigated the codes for stimulus and choice information first by examining activity in single neurons. In AC, we considered both stimulus and choice information, whereas, in PPC, we focused on choice information only, because PPC contained little pure stimulus information. Stimulus and choice information were small but significant in individual neurons, on average, and only a minority of neurons had large stimulus or choice information (Extended Data Fig. 4b–j). In both areas, individual neurons were briefly informative, with subsets of largely distinct neurons providing information at different time points in a trial^{10,22} (Fig. 2f–k). Single cells in AC and PPC were informative about the choice for ~ 100 and 300 ms, respectively, and individual AC cells were informative about the stimulus category for ~ 280 ms (calculated as a two-sided decay around the information peak of each cell; Fig. 2i, k). Therefore, in most individual neurons, information was weak and short-lived relative to the length of trials.

We thus reasoned that population codes may be important in AC and PPC, in particular if long-duration and diverse timescales were present. To examine the structure of functional interactions in population activity^{18,23–27}, we modified our encoding model to predict a given neuron's activity on the basis of the past activity of each of the other imaged neurons ('coupling predictors'; activity from up to ~ 2 s in the past, within defined lag ranges). These coupling predictors

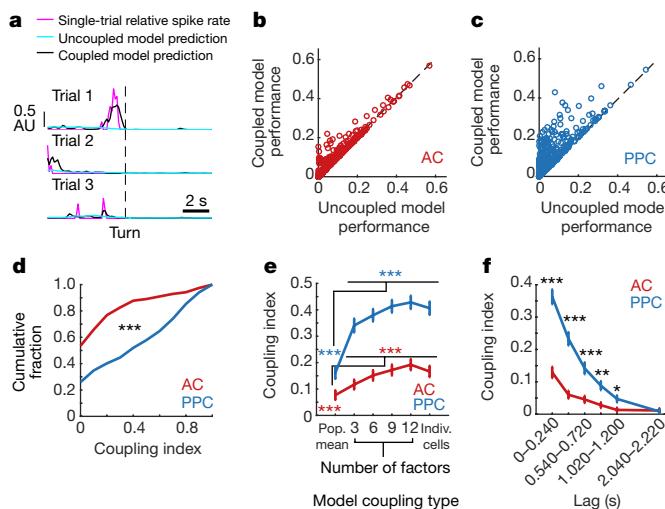


Figure 3 | PPC populations were more coupled than AC populations.

a, For a PPC neuron with high coupling index, single-trial responses (magenta) and predicted responses from the uncoupled (cyan) and coupled (black) models; AU, arbitrary units: relative spike rate or model prediction. b, c, Prediction performance of the coupled and uncoupled models for all AC and PPC neurons (circles). d, Cumulative distribution of the coupling index in AC (red) and PPC (blue) neurons. *** $P < 0.001$, Kolmogorov-Smirnov test. e, Coupling index in coupled model variants using as coupling predictors the mean population activity (Pop. mean), 3–12 factors extracted from the population activity, or all other simultaneously imaged neurons (Indiv. cells). *** $P < 0.001$, rank-sum test. f, Mean coupling index in AC (red) and PPC (blue) when coupling predictors were shifted by different temporal lags relative to the predicted neuron's activity. Data are mean \pm s.e.m. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; rank-sum test comparing PPC and AC. Only neurons with fraction explained deviance >0.1 in the coupled model were included in d–f ($n = 174/329$ AC neurons and $n = 185/386$ PPC neurons).

were included in a single model along with the task-related predictors described above²⁰ ('task predictors'; Fig. 3a–c). Here, 'coupling' indicates functional interactions between neurons without necessarily implying a direct causal connection between them. We quantified how strongly a neuron was coupled with other neurons in the local population by computing a 'coupling index': the difference in performance between the models with and without coupling predictors divided by the coupled model performance. Higher coupling indices indicated that other neurons' activity provided greater improvement in the prediction of a neuron's moment-to-moment activity patterns beyond what could be modelled with task features alone. Coupling is conceptually similar to 'noise' correlation²³, but it discounts, on a trial-by-trial basis, effects arising from shared tuning to measured task events and includes all other simultaneously imaged neurons.

Strikingly, the activity of PPC neurons had nearly three times more contribution from near-instantaneous coupling than did AC neurons, on average (coupling index: AC, 0.14 ± 0.02 ; PPC, 0.40 ± 0.02 ; $P < 0.001$, Kolmogorov-Smirnov test; Fig. 3d). Coupling between PPC neurons was present even at lags in activity up to ~ 1.25 s ($P < 0.001$, signed-rank test; Fig. 3f). In contrast, AC coupling was weak at all lags (Fig. 3f). This difference in coupling was confirmed by calculating partial Pearson's correlations for activity in neuron pairs (Extended Data Fig. 5). AC and PPC therefore had major differences in the structure of population activity, with higher coupling among neurons in PPC. In PPC, the time window of these interactions (>1 s) far exceeded the information timescale of single neurons (~ 0.3 s; Figs 2k and 3f).

We tested whether coupling was due to global population changes^{28,29} or to coordinated activity in small neuronal groups³⁰ by comparing the performance of variants of the coupled model. We compared coupled models that included as coupling predictors only the mean population activity or the activity of factorized subsets of cells. Global fluctuations

in activity contributed in part to coupling, but most coupling resulted from cell-to-cell coupling or coupling to activity in small subsets of cells (Fig. 3e).

The major difference in coupling between PPC and AC neurons was present outside the task context, including during passive listening to auditory stimuli and periods without stimulus presentation or running behaviour (Extended Data Fig. 5). Although we could not exclude potential contributions from unmeasured variables (for example, whisking, pupil diameter), the coupling difference, including at long lags, appeared unlikely to result solely from responses to task events (Extended Data Fig. 2o; see also Extended Data Fig. 6).

In a population of neurons with transient activity that tiles a task trial, as we observed in AC and PPC (Figs 1d, e and 2f–h), coupling between neurons could extend the coding timescale beyond what can be achieved with independent neurons, by combining individual cell responses in a population code. For example, when the coupling between two PPC neurons extends to lags comparable to the temporal offset between the neurons' activity, the resulting population signal lasts from the start of the first neuron's activity to the end of the second neuron's activity. The across-time activity dependencies revealed from time-lagged coupling suggest that stimulus or choice information signals should have consistency over the intervals at which these temporal dependencies occur. We therefore tested the hypothesis that coupling in a population might influence the temporal consistency of information encoded by the population.

To examine the temporal consistency of information, at each time point in a trial, we calculated the decoder posterior (that is, the continuous-valued prediction of either the stimulus category or choice at that time point on a single trial) and computed the correlation between posterior values at different time points in a trial, for all possible intervals between time points (Fig. 4a–d). A high correlation in posterior values across long time intervals indicated high temporal consistency in information. In AC, for both stimulus and choice, the correlation between posterior values dropped rapidly as a function of the interval between time points. In contrast, for choice in PPC, the correlation between posterior values decayed more slowly and was high at long time intervals (Fig. 4c–e). In PPC, information consistency remained high for several hundred milliseconds longer than in AC (Fig. 4e, Extended Data Fig. 7 and Supplementary Information). AC stimulus and choice signals thus fluctuated rapidly over time whereas PPC choice signals did not.

To quantify effects on information consistency at long timescales, such as those potentially related to time-lagged coupling, we fitted the temporal decay of the posterior correlation with an exponential curve with two time constants (τ_1 and τ_2) and focused on the longer time constant (τ_2). In agreement with higher temporal consistency and greater coupling at long time lags in PPC relative to AC, in PPC the long time constant had a larger magnitude and contributed more to the fitted curve than in AC (coefficient for τ_2 : AC, 0.35 ± 0.02 ; PPC, 0.63 ± 0.02 ; $P < 0.001$, z -test; Fig. 4f). Given the small contribution of long timescale effects in AC, going forward, we only considered information consistency in PPC choice signals.

We tested the relationship between coupling and information consistency by using an analysis approach to disrupt coupling while preserving single-neuron activity patterns. We shuffled trial identities for each neuron independently within each set of trials with the same stimulus and choice, and again computed the population decoder posterior. When coupling was disrupted, the consistency of the choice signal in PPC was shorter (Fig. 4f). Consistent with a role of coupling, information consistency increased with the size of the neuronal population when coupling was intact, but not when coupling was disrupted by shuffling (Supplementary Information and Extended Data Fig. 8). Even after disrupting coupling, differences in consistency remained between AC and PPC, probably because of single-neuron timescales (Fig. 2) and because shuffling cannot fully remove the effects of functional or anatomical coupling from single-neuron responses (Methods).

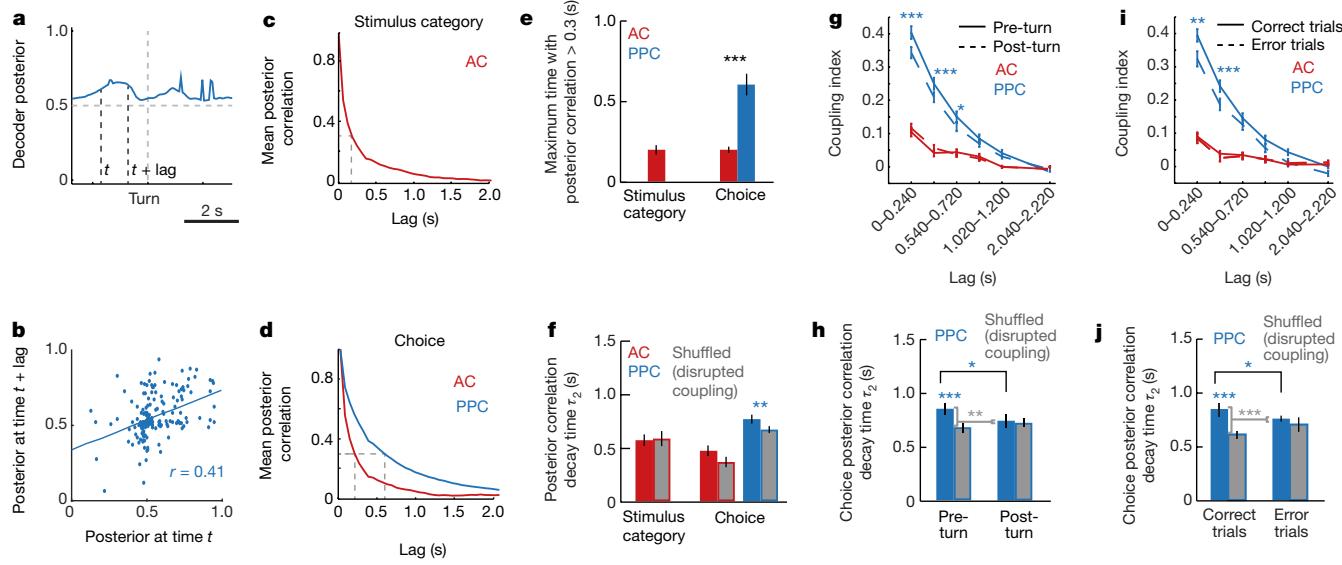


Figure 4 | Coupling is associated with a longer timescale of population codes for choice in PPC. **a**, Example population, instantaneous decoder posterior calculated on a single trial in PPC. **b**, Example demonstrating the calculation of posterior correlation. At times t and $t + \text{lag}$, the correlation coefficient between the posteriors at this interval was calculated for all trials (dots). **c, d**, Mean posterior correlation measured between all pairs of time points in the trial for stimulus and choice decoding. **e**, Time extent for which the mean autocorrelation functions in **c** and **d** is above 0.3 (dashed grey lines in **c** and **d**). *** $P < 0.001$, z -test. **f**, Time constant τ_2 of double exponential fits to posterior correlation functions. Coloured bars indicate unshuffled data; grey bars indicate data shuffled to disrupt coupling. The difference between coloured and grey bars

tests the contribution of coupling to the temporal consistency. Asterisks indicate significant differences between real and shuffled data, z -test from confidence intervals of fits. Error bars, 95% confidence intervals. **g**, Time-shifted coupling index, as in Fig. 3f, for pre- and post-turn periods. Asterisks indicate significant differences between pre-turn and post-turn data, rank-sum test. **h**, Time constant τ_2 of double exponential fits to posterior correlation time courses in pre-turn and post-turn data. Brackets show comparisons between the contribution of coupling (unshuffled – shuffled) with the consistency of the choice signal across conditions, z -test. Error bars, 95% confidence intervals. **i, j**, Same as **g** and **h**, for behaviourally correct versus error trials. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

Long timescales of information coding in PPC therefore resulted in part from population-level interactions, because, with coupling disrupted, information consistency was shorter by hundreds of milliseconds.

We examined whether the timescale of coding within a region had variability that was related to the behavioural context. We compared periods of the trial before and after the mouse executed its behavioural report as a left or right turn. Pre-turn choice information had the potential to be causal for the upcoming behavioural report, whereas post-turn choice information did not. In AC, coupling was similarly weak during the pre- and post-turn periods. In contrast, in PPC, coupling was higher and the temporal consistency of choice information was longer during the pre-turn period than during the post-turn period (Fig. 4g, h). In agreement with the difference in coupling between these two periods, disrupting coupling by trial shuffling had a significant effect on PPC choice consistency only in the pre-turn period (Fig. 4h).

We also tested whether the levels of coupling and consistency were related to the accuracy of behavioural choices. In AC, coupling was similarly weak during correct and error trials. In contrast, PPC coupling and choice information consistency were greater on correct trials than on error trials (Fig. 4i, j). Also, disrupting coupling with the trial shuffle shortened the timescale of choice consistency in correct trials but had little effect in error trials (Fig. 4j). PPC thus had strong coupling and temporally consistent choice signals on correct trials in the pre-turn period. On error trials and after a choice was reported, however, PPC had weaker coupling and consistency. Coupling and consistency therefore may be of importance for conveying signals relevant for accurate behaviour.

Together our results reveal that, despite short coding timescales in individual neurons, long timescales can emerge in neuronal populations. However, coding timescales were variable across cortex and depended on the structure of the population code. AC had relatively weak coupling and a short timescale (hundreds of milliseconds), which might aid representations of rapidly fluctuating stimuli and high dimensional sensory features. Previous studies have proposed that noise

correlations can have a detrimental, information-limiting effect^{23,26,29} and have thus suggested that sensory codes may benefit from weak coupling, which appears consistent with our findings in AC. However, in contrast, PPC had strong coupling and a long population timescale (~ 1 s), which appear to have a beneficial effect, because higher levels of coupling and temporal information consistency corresponded to more accurate task performance. In PPC, coupling timescales could be long enough to combine temporally separate inputs and could result in higher instantaneous information, because of information accumulation over time. We built a data-driven computational model that confirmed this effect of coupling (Supplementary Information and Extended Data Fig. 9). Further, from such a code, a downstream network could instantaneously read out a signal containing consistent and accumulated information about the recent estimate of the appropriate choice. Our model showed how such a temporally consistent choice signal could improve behavioural accuracy (Supplementary Information and Extended Data Fig. 9). We propose that codes underlying sensory representations and choice signals might differ substantially and that the structure of population codes may be a defining characteristic of cortical regions that contributes to the computations performed in each area.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 September 2016; accepted 13 June 2017.

Published online 19 July 2017.

- Hasson, U., Chen, J. & Honey, C. J. Hierarchical process memory: memory as an integral component of information processing. *Trends Cogn. Sci.* **19**, 304–313 (2015).
- Honey, C. J. et al. Slow cortical dynamics and the accumulation of information over long timescales. *Neuron* **76**, 423–434 (2012).
- Mauk, M. D. & Buonomano, D. V. The neural basis of temporal processing. *Annu. Rev. Neurosci.* **27**, 307–340 (2004).
- Murray, J. D. et al. A hierarchy of intrinsic timescales across primate cortex. *Nat. Neurosci.* **17**, 1661–1663 (2014).

5. Yang, Y. & Zador, A. M. Differences in sensitivity to neural timing among cortical areas. *J. Neurosci.* **32**, 15142–15147 (2012).
6. Buracas, G. T., Zador, A. M., DeWeese, M. R. & Albright, T. D. Efficient discrimination of temporal patterns by motion-sensitive neurons in primate visual cortex. *Neuron* **20**, 959–969 (1998).
7. Yang, Y., DeWeese, M. R., Otazu, G. H. & Zador, A. M. Millisecond-scale differences in neural activity in auditory cortex can drive decisions. *Nat. Neurosci.* **11**, 1262–1263 (2008).
8. Shadlen, M. N. & Newsome, W. T. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J. Neurophysiol.* **86**, 1916–1936 (2001).
9. Andersen, R. A. & Cui, H. Intention, action planning, and decision making in parietal-frontal circuits. *Neuron* **63**, 568–583 (2009).
10. Harvey, C. D., Coen, P. & Tank, D. W. Choice-specific sequences in parietal cortex during a virtual-navigation decision task. *Nature* **484**, 62–68 (2012).
11. Jenkins, W. M. & Merzenich, M. M. Role of cat primary auditory cortex for sound-localization behavior. *J. Neurophysiol.* **52**, 819–847 (1984).
12. Nakamura, K. Auditory spatial discriminatory and mnemonic neurons in rat posterior parietal cortex. *J. Neurophysiol.* **82**, 2503–2517 (1999).
13. McNaughton, B. L. *et al.* Cortical representation of motion during unrestrained spatial navigation in the rat. *Cereb. Cortex* **4**, 27–39 (1994).
14. Nitz, D. A. Tracking route progression in the posterior parietal cortex. *Neuron* **49**, 747–756 (2006).
15. Raposo, D., Kaufman, M. T. & Churchland, A. K. A category-free neural population supports evolving demands during decision-making. *Nat. Neurosci.* **17**, 1784–1792 (2014).
16. Whitlock, J. R., Pfuhls, G., Dagslott, N., Moser, M.-B. & Moser, E. I. Functional split between parietal and entorhinal cortices in the rat. *Neuron* **73**, 789–802 (2012).
17. Hanks, T. D. *et al.* Distinct relationships of parietal and prefrontal cortices to evidence accumulation. *Nature* **520**, 220–223 (2015).
18. Graf, A. B. & Andersen, R. A. Predicting oculomotor behaviour from correlated populations of posterior parietal neurons. *Nat. Commun.* **6**, 6024 (2015).
19. Park, I. M., Meister, M. L., Huk, A. C. & Pillow, J. W. Encoding and decoding in parietal cortex during sensorimotor decision-making. *Nat. Neurosci.* **17**, 1395–1403 (2014).
20. Pillow, J. W. *et al.* Spatio-temporal correlations and visual signalling in a complete neuronal population. *Nature* **454**, 995–999 (2008).
21. Panzeri, S., Harvey, C. D., Piasini, E., Latham, P. E. & Fellin, T. Cracking the neural code for sensory perception by combining statistics, intervention, and behavior. *Neuron* **93**, 491–507 (2017).
22. Hromádka, T. & Zador, A. M. Representations in auditory cortex. *Curr. Opin. Neurobiol.* **19**, 430–433 (2009).
23. Averbeck, B. B., Latham, P. E. & Pouget, A. Neural correlations, population coding and computation. *Nat. Rev. Neurosci.* **7**, 358–366 (2006).
24. Carnevale, F., de Lafuente, V., Romo, R. & Parga, N. An optimal decision population code that accounts for correlated variability unambiguously predicts a subject's choice. *Neuron* **80**, 1532–1543 (2013).
25. Harris, K. D. *et al.* How do neurons work together? Lessons from auditory cortex. *Hear. Res.* **271**, 37–53 (2011).
26. Cohen, M. R. & Maunsell, J. H. R. Attention improves performance primarily by reducing interneuronal correlations. *Nat. Neurosci.* **12**, 1594–1600 (2009).
27. Ecker, A. S. *et al.* Decorrelated neuronal firing in cortical microcircuits. *Science* **327**, 584–587 (2010).
28. Okun, M. *et al.* Diverse coupling of neurons to populations in sensory cortex. *Nature* **521**, 511–515 (2015).
29. Rabinowitz, N. C., Goris, R. L., Cohen, M. & Simoncelli, E. P. Attention stabilizes the shared gain of V4 populations. *eLife* **4**, e08998 (2015).
30. Bathellier, B., Ushakova, L. & Rumpel, S. Discrete neocortical dynamics predict behavioral categorization of sounds. *Neuron* **76**, 435–449 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Chettih and M. Minderer for developing the cell selection software, and J. Assad, J. Drugowitsch, O. Mazor, and members of our laboratories for comments on the manuscript. We also thank the Research Instrumentation Core at Harvard Medical School. This work was supported by a Burroughs-Wellcome Fund Career Award at the Scientific Interface, the Seale Scholars Program, the New York Stem Cell Foundation, the Armenian-Harvard Foundation, the Alfred P. Sloan Research Foundation, a NARSAD Brain and Behavior Research Young Investigator Award, National Institutes of Health grants from the National Institute of Mental Health BRAINS program (R01-MH107620) and the National Institute of Neurological Disorders and Stroke (R01-NS089521), the Autonomous Province of Trento (Grandi Progetti ATTEND), and the Fondazione Bertarelli. C.A.R. is a Life Sciences Research Foundation Simons Fellow. C.D.H. is a New York Stem Cell Foundation Robertson Neuroscience Investigator.

Author Contributions C.A.R. and C.D.H. conceived the project and designed the experiments. All authors contributed to the development of the concepts presented in the paper. C.A.R. performed the experiments. All authors designed the data analysis approaches. C.A.R. and C.D.H. conceived the application of the GLM and coupling approaches. E.P. and S.P. conceived information and consistency measures and modelling approaches. C.A.R. and E.P. performed the data analysis. C.A.R. and C.D.H. wrote the manuscript with contributions from E.P. and S.P. All authors contributed to the content and writing of the Supplementary Information.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations. Correspondence and requests for materials should be addressed to C.D.H. (harvey@hms.harvard.edu) or S.P. (stefano.panzeri@iit.it).

Reviewer Information *Nature* thanks J. Pillow and the other anonymous reviewer(s) for their contribution to the peer review of this work.

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

Mice. All experimental procedures were approved by the Harvard Medical School Institutional Animal Care and Use Committee and were performed in compliance with the Guide for Animal Care and Use of Laboratory Animals. Imaging data were collected from five male C57BL/6J mice (The Jackson Laboratory) that were ~ 7 weeks old at the initiation of behaviour task training.

Virtual reality system. We used a modified version of the visual virtual reality system that has been described previously³¹. Head-restrained mice ran on a spherical treadmill. Forward/backward translation in the maze was controlled by treadmill changes in pitch (relative to the mouse's body), and rotation in the virtual environment was controlled by roll of the treadmill (relative to the mouse's body). Images were back-projected onto a half-cylindrical screen (24-inch diameter) using a picoprojector (Microvision). Mazes were constructed using the Virtual Reality Mouse Engine (VirMEn³²) in MATLAB (Mathworks). Four electrostatic speakers (Tucker-Davis Technologies) were positioned in a semicircular array, centred on the mouse's head. The speakers were positioned at -90° , -30° , $+30^\circ$, and $+90^\circ$ in azimuth, with the speakers arranged from lateral to behind the mouse's head (Fig. 1a). Speakers were calibrated to deliver similar sound levels (~ 70 dB, varying randomly by ± 5 dB to further ensure that variations in sound level per se could not be used to complete the task) in a sound isolation chamber and at the location of the mouse's head using a random incidence microphone (PCB Piezotronics). External sounds were attenuated by lining the surfaces and surrounding walls of the behaviour/imaging apparatus with sound foam.

Behaviour task. Before behavioural training, dental cement was used to attach a one-sided titanium head plate to the skull of a 6- to 8-week-old mouse. Upon recovery, the mouse was put on a water schedule, receiving 0.8 ml of water in total per day. Body weight was monitored daily to ensure it was maintained above 80% of the pre-restriction measurement.

In the final version of the task that was used during imaging experiments, mice ran down the stem of the virtual T-maze, judged the location of sound stimuli to be either on the left or right, and reported decisions by turning left or right at the T-intersection for water rewards (4 μ l per reward). The use of an auditory cue dissociated the sensory information necessary for the perceptual decision from the ongoing visual cues needed to navigate through the virtual environment. The stem of the maze had grey walls with stripes to provide optic flow to the running mouse; black towers with white dots were positioned in the left and right arms, strictly as visual landmarks for navigation purposes. The sound stimuli were 1- to 2-s-long dynamic ripples³³ (broadband stimuli that were created by summing 32 tones with carrier frequencies spaced across 2–32 kHz, which each fluctuated at 10–20 Hz). The stimuli were designed to activate many neurons in AC, independently of the sound frequency tuning of individual neurons, as the timescale of sound frequency fluctuations was faster than the timescale of imaging frames (16 Hz). Three different ripples were used during the task. Each trial used only a single ripple. However, the different ripples did not result in distinguishable activity patterns ($P > 0.1$ using a support vector machine classifier to identify the ripple type on the basis of neuronal activity in AC). We therefore combined these trials together for analysis. The sound stimulus was activated when the mouse passed an invisible spatial threshold at ~ 10 cm into the T-stem and originated from one of eight possible locations. The stimulus was repeated after a 100 ms gap; repeats continued until the mouse reached the T-stem. Most trial durations allowed for three stimulus repeats. Naive mice were trained to first run down straight virtual corridors of increasing length for water rewards, while a 10 kHz tone 20 ms in duration was delivered at increasing pulsing frequency (2–10 Hz) as the mouse approached the end of the maze. In the second stage, mice learned to use the most laterally positioned left (-90°) or right ($+90^\circ$) cues (dynamic ripples, described above) to guide left or right turns in a T-maze. When performance exceeded 75% correct, additional sound locations were gradually added on a session-by-session basis, until all eight sound location conditions were included. Four sound locations corresponded to the locations of the four speakers, while four additional virtual sound locations at -60° , -15° , $+15^\circ$, and $+60^\circ$ were simulated using vector base intensity panning, where the same sound stimulus was delivered to two neighbouring speakers simultaneously, scaled by a gain factor³⁴. While the total sound level at the mouse's ear was calibrated for all eight locations, the sound level of the stimulus was changed randomly between trials, so that potential slight variations in sound level between locations could not be used as a cue to complete the task.

Sound location conditions were randomly selected, except that the most difficult sound location conditions ($\pm 15^\circ$) were presented half as often as the other sound location conditions. A 'reward tone' was played as the water reward was delivered on correct trials (when the mouse had reached ~ 10 cm into the correct arm of the T-maze), and a 'no reward tone' was played when the mouse reached ~ 10 cm into

the incorrect arm on error trials. The inter-trial interval was 3 s on correct trials and 5 s on error trials. Mice performed 200–300 trials in a typical session over approximately 45–60 min. Mice were able to perform the task interchangeably on two different behavioural set-ups with different sets of speakers, indicating that it was unlikely that mice used a non-location sound stimulus variable, such as differential timbre of the individual speakers, to perform the task.

Passive listening and 'no task/stimulus' activity contexts. After the behaviour task session, the virtual reality display was turned off, and imaging continued as sound stimuli were presented to the mouse. Stimulus sets included frequency-filtered and -shifted natural sounds (sourced from Cornell Laboratory of Ornithology), sinusoidally amplitude-modulated pure tones (10 Hz modulation)³⁵, and the location-dependent dynamic ripple stimulus set used during the task, all calibrated to sound levels at 70 dB. Stimuli were generated, attenuated, and delivered in MATLAB at 192 kHz sampling frequencies (National Instruments, PCI-6229). The same cells were imaged as in the behavioural task. Presentations of the sets of sinusoidally amplitude-modulated tones, natural sounds, and sound location stimuli were interleaved with 5-min periods of data acquisitions with no stimulus presentation ('no task/stimulus'). Mice mostly ran on the treadmill during this time, but also spent time resting.

Surgery. When mice reliably performed the full version of the task, the cranial window implant surgery was performed. Mice were given free access to water for 3 days before the surgery. Twelve to 24 h before the surgery, mice were given two doses of dexamethasone (2 μ g per g (body weight)). For the surgery, the mouse was anaesthetized with 1.5% isoflurane. The head plate was removed, and elliptical craniotomies were performed over AC and PPC on the left hemisphere (PPC centred at 2 mm posterior and 1.75 mm lateral to bregma; AC centred at 3.0 mm posterior and 4.3 mm lateral to bregma). A 10:1 viral mixture of tdTomato (AAV2/1-CAG-tdTomato) to GCaMP6 (AAV2/1-synapsin-1-GCaMP6f) or GCaMP6 alone was injected at three to six evenly spaced locations along the anterior-posterior axis of AC, and three injections spaced 200 μ m apart were made in the centre of PPC. A micromanipulator (Sutter, MP285) moved a glass pipette to ~ 250 μ m below the dura at each site, and a volume of approximately 30 nl was pressure-injected over 5–10 min. Dental cement sealed a glass coverslip (3 mm diameter) over a drop of Kwik Sil (World Precision Instruments) on the craniotomy, and a new head plate was implanted, along with a rubber ring, to interface with a black rubber objective collar to isolate fluorescence photons from those generated by the virtual reality system. Mice recovered for 2–3 days after surgery before being placed back on the water schedule. Imaging was performed daily in each mouse, starting 4–6 weeks after surgery and continued for 4–12 weeks.

Two-photon imaging. Images were acquired using a two-photon microscope (Sutter MOM) at 15.6 Hz frame rate and 256 pixel \times 64 pixel resolution (~ 250 μ m \times 100 μ m) through a $\times 40$ magnification water immersion lens (Olympus, numerical aperture 0.8). On alternating days, either AC or PPC was imaged, at a depth of 150–300 μ m, corresponding to layers 2/3. For AC imaging, the objective was rotated to approximately 35–40° from vertical, and for PPC imaging, it was rotated to about 5–10° from vertical. Each field of view contained approximately 40–70 neurons. A Ti-sapphire laser (Coherent) tuned to 920 nm delivered excitation light. Emitted light was isolated using a dichroic mirror (562 nm long-pass) and green (525/50 nm) and red (609/57 nm) band-pass filters (Semrock). ScanImage (version 3, Vidrio Technologies) was used to control the microscope. Outputs controlling the galvanometers and the audio speakers, along with an iteration counter from VirMEn, were collected by a digital interface (Digidata, Molecular Devices), allowing offline alignment of imaging frames to behaviour events.

Data processing. Imaging datasets from seven AC fields of view and seven PPC fields of view were included from five mice. Movies of imaging frames collected during the task, passive listening, and 'no task/stimulus' activity contexts were concatenated for motion correction, cell body identification, and deconvolution. In brief, after motion correction³⁶, correlations in fluorescence activity time series between pixels within ~ 60 μ m were calculated. Fluorescence sources (putative cells) were identified by applying a continuous-valued eigenvector-based approximation of the normalized cuts objective to the correlation matrix, followed by *k*-means clustering, yielding binary masks for all identifiable fluorescence sources. Only datasets with at least 35 cells were included for further analysis. To estimate potential neuropil contamination, we regressed the cell body fluorescence signal against signal from surrounding pixels during time points when the cell of interest was not active and used a robust-regression algorithm, and then removed neuropil contamination during the $(F - F_{\text{baseline}})/F_{\text{baseline}}$ calculation by subtracting a scaled version of the neuropil signal from the cell body signal. F_{baseline} was the eighth percentile spanning 500 frames (~ 30 s) around each frame. MATLAB scripts implementing these custom algorithms are available online (https://github.com/HarveyLab/Acquisition2P_class.git) or upon request from the corresponding authors. Fluorescence traces were deconvolved to estimate the relative spike rate

in each imaging frame³⁷, and all subsequent analyses were performed on the estimated relative spike rate to reduce the effects of GCaMP6f signal decay kinetics. Because this estimate is the spike rate relative to baseline activity, without perfect knowledge of the fluorescence change associated with a single spike, we report the estimated relative spike rate in arbitrary units (J. Vogelstein, personal communication).

For visualization and for decoding analyses, data were temporally aligned either to the sound onset (for stimulus-related analyses) or to the moment of the turn, defined by when the mouse entered the short arm of the maze (for choice-related analyses).

Data inclusion criteria. To be included for further analysis, mouse performance during the imaging session had to exceed 65% correct, to ensure that mice were performing the task well above chance (50%), but had to be lower than 80% correct, as our analyses required a significant number of error trials to decouple stimulus and choice. Each field of view was required to contain at least 35 identified cells. Coupling was estimated in datasets with fewer than 35 neurons (not shown), and results were similar to those reported here. Finally, AC fields of view had to contain sound-frequency-selective neurons with preferences in the sub-ultrasonic range for further analysis, to ensure that we were imaging in tonotopic AC.

GLM (encoding model). Our encoding model extended the approach taken in refs 19, 20 to calcium imaging data recorded in populations of neurons in a behaving animal. It allowed us to model, for single neurons, the time-dependent effects of all measured variables related to the task and the animal's behaviour simultaneously on neuronal activity during single trials. Simpler approaches regressing the spike rates of individual neurons against the values of only a few variables of interest (for example, choice or stimulus only) have been useful in tasks involving fewer variables and with task timing determined by the experimenter. However, in our case, mice controlled trial timing with their own running speed and could change their view of the virtual reality environment by running laterally. We therefore took the approach of trying to explain trial-trial variability due to differences in the timing of stimuli or differences in behavioural variables such as running speed across trials. Here, our model took into account many features describing the mouse's sensory environment (auditory and visual) and behavioural actions (such as running movements).

We used a Bernoulli GLM to weight task variables (task predictors, uncoupled model; Fig. 2a) or task variables and activity of other neurons (coupling predictors, coupled model; Fig. 2a) in predicting each neuron's binarized activity (time series of relative spike rates were converted to vectors of ones and zeros: different thresholds for activity were compared, but did not change the results, so all non-zero relative spike rates were set to one). The Bernoulli model was primarily selected as, owing to the sparseness of the recorded activity, the most prominent feature of the data was whether any signal was present at a given time. Poisson and multinomial versions of the GLM were also built and tested, and yielded qualitatively similar fits and information estimates.

Uncoupled model predictors. Task-related predictors, measured at higher time resolution than the imaging, were binned by averaging as necessary to match the sampling rate of imaging frames (15.6 Hz). The behaviour variables included the running velocity on the pitch and roll axes of the treadmill (relative to the mouse's body axis), *x* and *y* position in the virtual maze, onset times and locations of sound stimuli, the mouse's virtual view angle in the maze, turn direction (choice), and reward and error signal delivery times (Extended Data Fig. 2). Sound stimulus and reward/error events were represented as boxcar functions that were set to a value of one at the time of onset and zero everywhere else. Predictors were convolved with behaviourally appropriate sets of basis functions (evenly spaced Gaussian kernels), to produce the task predictors (Extended Data Fig. 2b). This allowed us to fit time-dependent modulation of neuron responses by the task predictors, as follows.

For sound stimulus onsets at each of the 8 possible sound locations, 12 evenly spaced Gaussian basis functions (170 ms half-width at half-height) extended 2 s forward in time from each sound onset. First, second, and third repeats were represented separately, because of the adaptation-related effects we report in AC neuron responses during the task and passive listening contexts (Extended Data Fig. 1). This resulted in 12 basis functions per repeat per sound location \times 3 repeats \times 8 locations for 288 sound predictors. Reward and error delivery times were convolved on separate channels with four Gaussians (500 ms half-width at half-height), and extended 2 s forward in time from reward/error signal onset times (eight basis functions in total for reward and error). Examples of the types of basis function used for events such as sound and reward delivery are shown in Extended Data Fig. 2c.

Running velocity measurements were separated into four channels, (1) forward, (2) reverse, (3) left, and (4) right directions, on the basis of rotations of the treadmill about the pitch and roll axes, relative to the mouse's body axis. Changes in running speed could be either responded to or anticipated by neuronal activity, so time series of running velocity were convolved with four evenly spaced Gaussian basis

functions (240 ms half-width at half-height) extending 1 s both forwards and backwards in time (8 bases total for each running direction: forward, reverse, left, and right; Extended Data Fig. 2b; 32 basis functions in total for running predictors). Changes in virtual reality view angle were modelled similarly, by two channels: (1) leftward and (2) rightward directions. Each channel was convolved with three evenly spaced Gaussian basis functions extending in the forward and reverse time directions (half-width at half-height, 320 ms; 12 total basis functions for view angle).

Upcoming left and right choices were represented on separate channels by two types of basis function: (1) 30 evenly spaced Gaussian 'place fields' (half-width at half-height: 1/10 of the total maze length) along the stem of the T-maze (Extended Data Fig. 2d) preceding the turn, and (2) by 9 evenly spaced Gaussian temporal basis functions extending 4 s forward in time (440 ms half-width at half-height) from the turn (behavioural choice), defined by when the mouse turned into the short arm of the maze. In total, left and right turns were convolved with 78 basis functions.

For convenience, all predictors were normalized to their maximum values before being fed into the model. The total number of predictors in the uncoupled model was 419 (420 if counting the constant predictor which corresponded to the average activation probability of each individual cell). The basis functions for each behavioural variable were selected to optimally predict responses of simulated neurons with simple tuning to that variable, with response properties similar to those that we observed in our AC and PPC datasets. We optimized the sound onset predictors by simulating neurons with sound-location-selective responses with different response latencies, as we observed in AC (Extended Data Fig. 1). For example, we defined a neuron that would respond only when a sound was played from the -90° location with a latency of 100 ms, on the first sound repeat, and used the actual behaviour data collected during imaging experiments to simulate the timing and response magnitude of such a neuron, with added Poisson noise. The number, width, and spacing of basis functions were then systematically varied until the model performance was maximized in predicting this simulated neuron's response. We repeated this procedure for neurons responding with higher magnitude to later stimulus repeats (like the neuron in Extended Data Fig. 1h), with more broad sound location tuning (that is, responding to all sound locations, to only locations 0 to $+90^\circ$, or to -45 to $+45^\circ$), and for neurons with a diversity of sound onset/offset latency responses. Furthermore, we examined the beta coefficients in these model fits to ensure that weights were not aberrantly assigned to non-sound-related predictors.

We optimized the basis functions for other variables using similar methods, simulating neurons selective for running speed, spatial position in the maze, reward or error timing, view angle, and turn direction (behavioural choice). We found that combining the predictors for upcoming turn direction and maze position was the optimal solution, producing two sets of 'place fields' for upcoming left and right turns. These predictors were able to account for the type of response we observed in PPC neurons (choice-selective neurons tended to respond at particular maze positions rather than at particular times in the trial). Furthermore, we observed that these spatial turn predictors prevented contamination of sound-selective and turn-selective predictor weights in the model fits.

Coupled model predictors. To compare the dependence of each neuron's activity on external behavioural correlates (uncoupled model) and the activity of other neurons in the local population, we developed a coupled version of the model, where the activity of other neurons was also included as predictors in the GLM (coupled predictors; Fig. 2a). By comparing the additional fraction of explained deviance^{38,39} (see below) beyond the uncoupled model's fraction of explained deviance, we could quantify the level of coupling of each neuron with other neurons, while also taking into account behavioural correlates that might commonly drive neurons within the population. This comparison is related to the measure of noise correlations, which attempts to measure the correlated variability of responses after subtracting the averaged response to repeated stimulus presentations ('signal correlations'). The advantage of this GLM approach is that, on a single-trial basis, our model accounts for variability in the timing and magnitude of behavioural- and stimulus-related predictors simultaneously. Furthermore, by including all simultaneously imaged neurons, the model can reveal group-wise interactions, which is not possible with pairwise noise correlation measurements²⁰.

For a given neuron being fitted, the relative spike rate of each other neuron and the population mean (excluding the cell being fitted) was convolved with two boxcar functions extending \sim 120 ms forward in time from the activity of predictor neurons (each boxcar was non-zero for only a single imaging frame, or \sim 60 ms). To test for the presence of coupling across longer lags, four evenly spaced coupling basis functions (as above, each a boxcar function that was non-zero for a single imaging frame) were built at lags that were shifted successively further from zero for different versions of the model (Fig. 3f).

Versions of the coupled model that included only the population mean, factorizations of the population activity selected via non-negative matrix factorization, or the full population of individual neurons were compared (Fig. 3e). Non-negative matrix factorization was preferred over other dimensionality reduction techniques, such as principal component analysis, as it provided a decomposition that could be naturally interpreted as a sum of parts⁴⁰—in this case, contributions from partly overlapping neuronal subpopulations. Non-negative matrix factorization was computed by running the ‘nmf’ MATLAB function on the full-time series of neurons in the population, excluding the neuron being fitted.

GLM fitting procedure. All predictors were maximum-normalized, for convenience, and z-scored before the fitting procedure.

We fitted the GLMs to each single neuron’s activity individually, using the *glmnet* package in R⁴¹ with elastic-net regularization, which smoothly interpolated between L_1 and L_2 type regularization according to the value of an interpolation parameter α , such that $\alpha = 0$ corresponded to L_2 and $\alpha = 1$ to L_1 . We selected $\alpha = 0.95$, allowing for a relatively small number of useful predictors to be selected by the model out of many potentially correlated predictors as in pure L_1 regularization, while at the same time avoiding issues with degeneracies that could arise if the correlations between predictors were very strong⁴¹. Additional penalty factors ($\times 10$) were applied to the coupling predictors to reduce the number of selected coupling predictors to the smallest number necessary to increase model performance²⁰. The value of the shrinkage parameter for the elastic-net was chosen by threefold cross-validation on the training data.

Within the training dataset (70% of trials), cross-validation folds were pre-selected so that trials with specific combinations of sound locations and choices were evenly divided among them. For instance, if there were 30 trials containing sound location +90° and left choice by the animal, 10 trials were randomly selected for each cross-validation fold. The test dataset (30% of trials), also containing a similar distribution of trial conditions, was left out of the fitting procedure entirely, and was used only for testing the model performance. Each model was thus fitted and tested on entirely separate data, removing over-fitting concerns. This train and test procedure was repeated ten times, with random subsamples of the data included in train and test segments.

GLM model performance. Model performance was quantified by computing the fraction of explained deviance³⁹ of the model. In addition to the full coupled and uncoupled models, we also fitted a null model to each cell’s activity. In the null model, only a constant (single parameter) was used to fit the neuron’s activity and no time-varying behaviour or coupling predictors were included as predictors. We calculated the deviance of the null, uncoupled, and coupled models; then, for the coupled and uncoupled models, we calculated the fraction of null model deviance explained by the model ((null deviance – model deviance)/null deviance). Deviance calculations were performed on a test dataset (30% of the data), which had not been included in the fitting procedure, and this train/test procedure was repeated ten times on randomly subsampled segments of the data.

By comparing the model performance (fraction of explained deviance) in the coupled model with the performance of the uncoupled model, we could estimate the level of correlation between a given neuron and the neurons in the simultaneously imaged population. Only neurons for which the overall GLM performance gave a fraction of deviance explained above 0.1 were included in this analysis, to avoid the lowest quality fits (the major results were present regardless of the threshold applied). The ‘coupling index’ (CI) compared the improvement in model performance when adding coupling, for each neuron:

$$\text{CI} = (d_c - d_u)/d_c \quad (1)$$

where d_c is the fraction of deviance explained in the coupled model and d_u is the fraction of deviance explained in the uncoupled model (Fig. 3d). Although the coupled model included more predictors than the uncoupled model, over-fitting was prevented by testing the model on the 30% of data points (test set) that were completely left out of the threefold cross-validated fitting procedure. A coupling index of 1.0 indicated that all of the explained deviance came from the coupled predictors, while values approaching 0 indicated that more deviance was explained by the uncoupled (behaviour) predictors. The coupling index was calculated for each of the ten random train/test subsamples, and the average across these ten calculations is reported for each individual neuron.

GLM model performance in passive listening and ‘no task/stimulus’ activity contexts. To help exclude the possibility that coupling parameters simply allowed the model to explain behavioural correlates not included in the uncoupled model, rather than neuron–neuron correlations, we measured coupling in other behavioural contexts, where behavioural variables and external stimuli were either not present or organized differently (Extended Data Fig. 5). In brief, the model was trained and tested on 70% and 30% of data, respectively, and this subsample was repeated ten times on random segments of the data within each context. The coupling index was calculated as above.

Cell-cell model: GLM model performance with no behavioural predictors. We used an additional approach to test whether the GLM might have misattributed correlates due to common drive from behavioural variables to the coupling predictors, where measured coupling might have reflected signal correlations either to modelled or to unmodelled behavioural variables. To estimate the upper limit U to which this might have occurred, we compared the performance of a version of the GLM that included only the coupling predictors (cell–cell model) with the improvement in model performance achieved by adding the coupling predictors to the uncoupled model:

$$U = d_{\text{cxc}} - (d_c - d_u) \quad (2)$$

where d_{cxc} is the fraction of deviance explained of the cell–cell model, d_c is the fraction of deviance explained of the coupled model, and d_u is the fraction of deviance explained of the uncoupled model (Fig. 3b, c and Extended Data Fig. 6). We reasoned that in the cell–cell model the behavioural variables that we knew explained AC and PPC activity well could bleed through into the coupling predictors, because of the neurons’ signal correlations. In this case, the fraction of deviance explained for the cell–cell model was expected to be much larger than the coupling value obtained from comparing the coupled and uncoupled models, because the cell–cell model’s value would include both coupling and behavioural variable bleed-through. In contrast, if the cell–cell model’s performance was similar to the value of coupling from the coupled–uncoupled model comparison, then it was unlikely that the behavioural variables could bleed-through into the coupling predictors.

We compared the distribution of coupling ($d_c - d_u$) measured in AC and PPC with the difference in the cell–cell model performance and coupling (U ; Extended Data Fig. 6c), finding that coupling in AC neurons was restricted to values below this estimate of the upper bound on coupling explainable by task-related variables, while coupling in many PPC neurons exceeded this upper bound. We concluded from this analysis that not all coupling in PPC neurons could be explained by common drive by the measured behavioural variables. In addition, we considered this estimate to be a likely large overestimate of the contribution from unmodelled behavioural variables, because the modelled behavioural features (sound stimuli, choice, running patterns) were expected to be the best behavioural correlates of AC and PPC activity.

Partial Pearson’s correlation. Partial Pearson’s correlations were computed between all pairs of simultaneously recorded neurons as follows. For each pair of neurons, and for all available time-point pairs (each time point corresponding to one imaging frame, or ~60 ms) within trials of the same stimulus category and choice condition, the partial Pearson’s correlation between the activity of the neurons was computed, discounting the effect of lateral running speed (MATLAB function ‘partialcorr’). Time-point pairs were then sorted by their difference (lag), and partial correlations for those with the same lag were averaged together. For each lag, the four partial correlations measured for each neuron pair (two stimulus categories by two choices) were then averaged, weighting by the number of trials in each condition.

In the passive and ‘no task/stimulus’ behavioural contexts, partial Pearson’s correlations were computed as follows. For each pair of simultaneously recorded neurons and for each trial, the partial Pearson’s correlation between the activity of the neurons within that trial was computed, discounting the effect of the lateral running speed. The partial correlations were then averaged across trials, yielding the final estimate for the partial Pearson’s correlation of the neuron pair. Since, by experimental design, each stimulus location in the passive context had the same number of trials and each trial had the same length, this was equivalent to computing the partial correlations separately for each stimulus location or category and then averaging across locations or categories. Note also that the ‘no task/stimulus’ behavioural context only had one trial for each experimental session.

In the ‘no task/stimulus’ context, we also computed Pearson’s correlations between pairs of simultaneously recorded neurons during periods where the mouse was stationary on the ball. In this case, correlations were computed as above, but without discounting running speed.

Decoder. To estimate the information represented in AC and PPC about the sound stimulus or the behavioural choice, we built a decoder that used the recorded cell responses to estimate the probability of each stimulus or choice condition on single trials. Our encoding model (GLM, see above) lent itself naturally to act as the core of the decoder (Fig. 2a). We decoded either stimulus or choice from single-trial population activity (population decoder), or from single-trial activity of individual neurons (single-cell decoder), by computing the probability of external variables given population or single-neuron activity using Bayes’ theorem and population or single-neuron response probabilities estimated through the uncoupled GLM and its predictors in that particular trial. Note that, because the encoding model took all measured behavioural variables into account, these predicted responses reflected trial-to-trial differences in running speed or virtual reality view angle in

addition to the stimulus and choice conditions. However, since we decoded only sound category or choice, we integrated away the effect of all behavioural variables other than choice and stimulus category, as detailed in the following paragraphs.

We first explain the instantaneous decoder that predicts stimulus or choice on the basis of the observation of an instantaneous population response ('instantaneous' meaning during one imaging frame, ~ 60 ms) $\underline{r}(t) = \{r_1(t), \dots, r_N(t)\}$ made of the activation $r_i(t)$ of cell i ($i = 1, \dots, n$) at time frame t . Call the variable to be decoded v , with $v = \text{left}, \text{right}$ meaning either presented stimulus (sound in the left or right part of the space) or behavioural choice (left or right choice, respectively). Call x the time courses of the GLM predictors that, by design, completely specify v , such that v can be thought of as a simple function $v = v(x)$: if v is choice, x is the time course of the 'upcoming left and right choice' predictors; and if v is presented stimulus, x is the time course of the indicator functions representing the presence of sound in the eight speakers (see Extended Data Fig. 2). Call \tilde{x} the time courses of all the other task variables (for example, running speed). With this convention, the probability of observing a neural activation $r_i(t)$ of cell i at time frame t can be compactly written as $p_i(r_i(t)|x, \tilde{x})$. For each trial in the training set, we computed all probabilities $p_i(r_i(t)|x, \tilde{x})$ from the predictors of the uncoupled GLM in that trial. Assuming cell activities to be conditionally independent given the predictors, we then computed the probability of observing a population activity $\underline{r}(t) = \{r_1(t), \dots, r_N(t)\}$ as follows:

$$p(\underline{r}(t)|x, \tilde{x}) = \prod_{n=1}^N p_n(r_n(t)|x, \tilde{x}) \quad (3)$$

The instantaneous Bayesian decoder then inverted this probabilistic model using Bayes' theorem to compute the posterior probability $p(x|\underline{r}(t))$ of observing each possible value of x by multiplying $p(\underline{r}(t)|x, \tilde{x})$ by the prior probability $p(x, \tilde{x})$ of each combination of behavioural variables and then integrating over the dummy variables \tilde{x} that we did not use in decoding, as follows:

$$p(x|\underline{r}(t)) = \sum_{\tilde{x}} p(x, \tilde{x}|\underline{r}(t)) \propto \sum_{\tilde{x}} p(\underline{r}(t)|x, \tilde{x})p(x, \tilde{x}) \quad (4)$$

where the prior distribution $p(x, \tilde{x})$ was taken to assign equal probability to each instance (x, \tilde{x}) that was observed in the training data and was zero for any other combination of behavioural variables that was never observed in the training data. The above integration over dummy variables is suitable, even taking into account the possibly sparse sampling of the dummy variables owing to the limited number of trials, for our principal purpose, which was to ensure that these variables could not contribute spurious information about the variable x when decoding the most likely value of x from equation (4). The decoded variable \hat{v} at time t was finally selected as the one whose corresponding values of x had the maximum posterior probability (maximum a posteriori decoding):

$$\begin{aligned} \hat{v}(\underline{r}(t); t) &:= \arg \max_{v'} \sum_{x: v(x)=v'} p(x|\underline{r}(t)) \\ &= \arg \max_{v'} \sum_{x: v(x)=v'} \sum_{\tilde{x}} p(\underline{r}(t)|x, \tilde{x})p(x, \tilde{x}) \end{aligned} \quad (5)$$

To make a concrete example, to decode the stimulus presented to the animal in a given trial at time t we partitioned the training data in trials where the sound was on the left (sound locations number 1, 2, 3, and 4; that is, all those instances of x such that $v(x) = \text{left}$) and trials where the sound was on the right (sound locations number 5, 6, 7, and 8, corresponding to $v(x) = \text{right}$). We then computed the likelihood of the observed activity $\underline{r}(t)$ with respect to the time courses of the predictors in all training trials, and summed the likelihood for each trial set (stimulus = left and stimulus = right). We then compared the two values obtained, and decoded left if the value for the subset of trials with stimulus on the left was the highest, and right otherwise.

The cumulative decoder was defined in an analogous fashion, but operated (rather than on the instantaneous responses and their probability as in equation (1)) on the time ensemble of single-trial population responses $\underline{r}(1), \dots, \underline{r}(t)$ computed on a whole set of time frames from a starting time frame 1 to time frame t . The probability $p_c(\underline{r}(1), \dots, \underline{r}(t)|x, \tilde{x})$ of the single-trial population responses in this time ensemble was computed assuming conditional independence of cell activity across time given the behavioural variables:

$$p_c(\underline{r}(1), \dots, \underline{r}(t)|x, \tilde{x}) = \prod_{\tau=1}^t p(\underline{r}(\tau)|x, \tilde{x}) = \prod_{\tau=1}^t \prod_{n=1}^N p_n(r_n(\tau)|x, \tilde{x}) \quad (6)$$

We note that we tried very extensively to decode using variations of GLM models that included coupling parameters between cells²⁰, but these more parameter-rich probability models did not increase the amount of information decoded about

choice or stimulus in AC or PPC, even though the coupled encoding model explained more trial-to-trial variability than the uncoupled model (Fig. 3). The results presented here can be thus considered as fully cross-validated lower bounds to the total information about stimulus or choice that could be extracted from neural responses, and (although we could not find a better performing decoding model) we cannot exclude the possibility that more information could be extracted with more refined models or if more data were available for training.

Estimation of information about stimulus or choice. For each imaging session, the experimental data were randomly split into equally sized training and testing datasets, ensuring stratification of sound location and choice combinations. The GLM was fitted on the training set as detailed above (see section on 'GLM fitting procedure'). All data were temporally aligned to the sound onset (for stimulus decoding) or the moment of the turn, defined by when the mouse entered the short arm of the maze (for choice decoding). Stimulus and choice were decorrelated in training and testing data by randomly subsampling the available trials so that each combination of stimulus and choice (that is, left stimulus, left choice; left stimulus, right choice; etc.) was equally represented. There were on average 7.2 ± 2.3 (mean \pm s.e.m.) trials per condition, with a minimum of 4 trials for any condition, and this random subsampling was repeated 20 times. This decorrelation ensured that we could separate pure information about choice from pure information about stimulus. This separation was useful to ensure that any difference in the resulting timescales of these signals would not be cancelled out by the possible mixing of these two signals. The GLM probability model was used to decode stimulus or choice on the test dataset using the Bayesian decoder outlined above. Only a randomly selected subpopulation of 37 cells was used for decoding to control for the size variability of the populations recorded, as this was the minimum population size across imaging sessions. The information about stimulus or choice decoded from neural population activity was computed as the mutual information between the real value v of the variable and the one \hat{v} decoded from neural activity, as follows^{42,43}:

$$I = \sum_{v, \hat{v}} p(v, \hat{v}) \log \frac{p(v, \hat{v})}{p(v)p(\hat{v})} \quad (7)$$

The limited-sampling bias in the information estimate was corrected for by subtracting the analytical estimate of the bias⁴⁴. This splitting–subsampling–decoding procedure was repeated 9,600 times (10 train/test splits, 20 subsamplings for each train/test split, 48 subsamplings of the neural population), and the information estimates were averaged together to yield the final result for each imaging session.

Overall, this method allowed us to quantify information on a meaningful scale (in bits), enabling comparisons between different decoders (for example, instantaneous versus cumulative). Furthermore, the decoding framework provided a natural way of relating the instantaneous and the cumulative decoders, and of bridging single-cell and population levels, by the conditional independence assumptions in equations (3) and (6). We note that all this would have not been the case with other approaches based on the encoding model only, such as an analysis of the distribution of the fitted GLM parameters.

Shuffling procedure to disrupt neuron–neuron correlations. To assess the effects of coupling between neurons on information coding and timescales, we shuffled trials using a method that disrupted functional coupling while maintaining activity time courses in individual cells. Within subsets of trials of the same behavioural choice and stimulus (for example, sound location = -90° and left choice, or sound location = -90° and right choice, for the data in Fig. 4), we shuffled the identities of trials for each neuron independently. In practice, shuffling trial identities meant shuffling single-neuron recorded activities across trials in the testing set and single-neuron activation probabilities $p_i(r_i(t)|x, \tilde{x})$ (see section on 'Decoder') across trials in the training set. Thus, on average, individual neurons' responses to the stimulus or choice condition were maintained (thus signal correlations were maintained), while coupling (noise correlations) among neurons was disrupted (and note that this did not affect the correlations between the predictors x, \tilde{x} from the point of view of each single-cell GLM). Importantly, the single-neuron autocorrelations were maintained in the shuffled condition (as shuffling did not alter the time course of single-cell activities or predictions), so the contribution of coupling could be distinguished from the contribution of single-neuron timescales and the timing of the activity of single neurons across time to information coding.

The shuffling procedure could not remove all effects of coupling. The shuffle effectively removed across-trial covariations in activity between neurons. However, because the shuffle was performed in analysis, rather than through an experimental disruption of correlations, it is likely that some effects of coupling remained in each single neuron's recorded activity. For example, imagine the case in which there is a transient external input to a network at time $t=0$. In the absence of coupling, all cells will stop responding within a short time (for example, by $t=1$) depending on the single-neuron timescale. However, in the presence of coupling, the population

can follow an informative, seconds-long trajectory with different cells active at different times. As a result of coupling, each neuron could potentially be active after $t=1$. The shuffle would be unable to remove this extended timescale that is present in the measured activity of individual neurons because of coupling. Shuffling therefore is effective at removing patterns of coupling in the population, but it does not modify any individual cell's activity and thus cannot remove all effects of coupling. Our disruption of coupling by shuffling is therefore probably an underestimate of how much population interactions contribute to the timescales of coding.

Unlike the stimulus location category or choice, the time courses of other task predictors could vary differently from trial to trial. To ensure that disruptions of the effects of task predictors in the trial shuffling procedure could not account for the differences in consistency or population activity dynamics between real and shuffled data, we compared the variability of running speed and maze position for each time point across the trial between correct and error trials and time points pre- and post-choice. The variability did not differ between trial types, but was significantly higher post-choice than pre-choice ($P < 0.001$; rank-sum test); thus, the effects of shuffling trials on consistency were not due to disruptions in task predictors, as we would have expected a greater effect of shuffling on consistency post-choice if that were the case (Fig. 4h).

Decoder posterior consistency. For each experimental trial, the instantaneous posterior probability $p(v'|z(t)) = \sum_{x:v(x)=v} p(x|z(t))$ was computed as described above for the real value of the variable to be decoded (for example, $p(\text{left}|z(t)) = \sum_{x:v(x)=\text{left}} p(x|z(t))$) if decoding choice on a trial where the animal chose to go left) and averaged across all train/test splits and subsamplings. To capture the consistency over time of the decoder posterior, the Pearson's correlation coefficient between the average posterior probabilities at all available time-point pairs was computed. Time-point pairs were then sorted by their difference (lag), and posterior correlations for the pairs with the same lag were finally averaged together to yield the consistency measure for each imaging session at each time point. All sessions were then averaged together, giving the values shown in Fig. 4c, d. For the analysis of shuffled data consistency, the activity of each cell was shuffled randomly among all trials with the same sound location and the same choice (see section on 'Shuffling procedure').

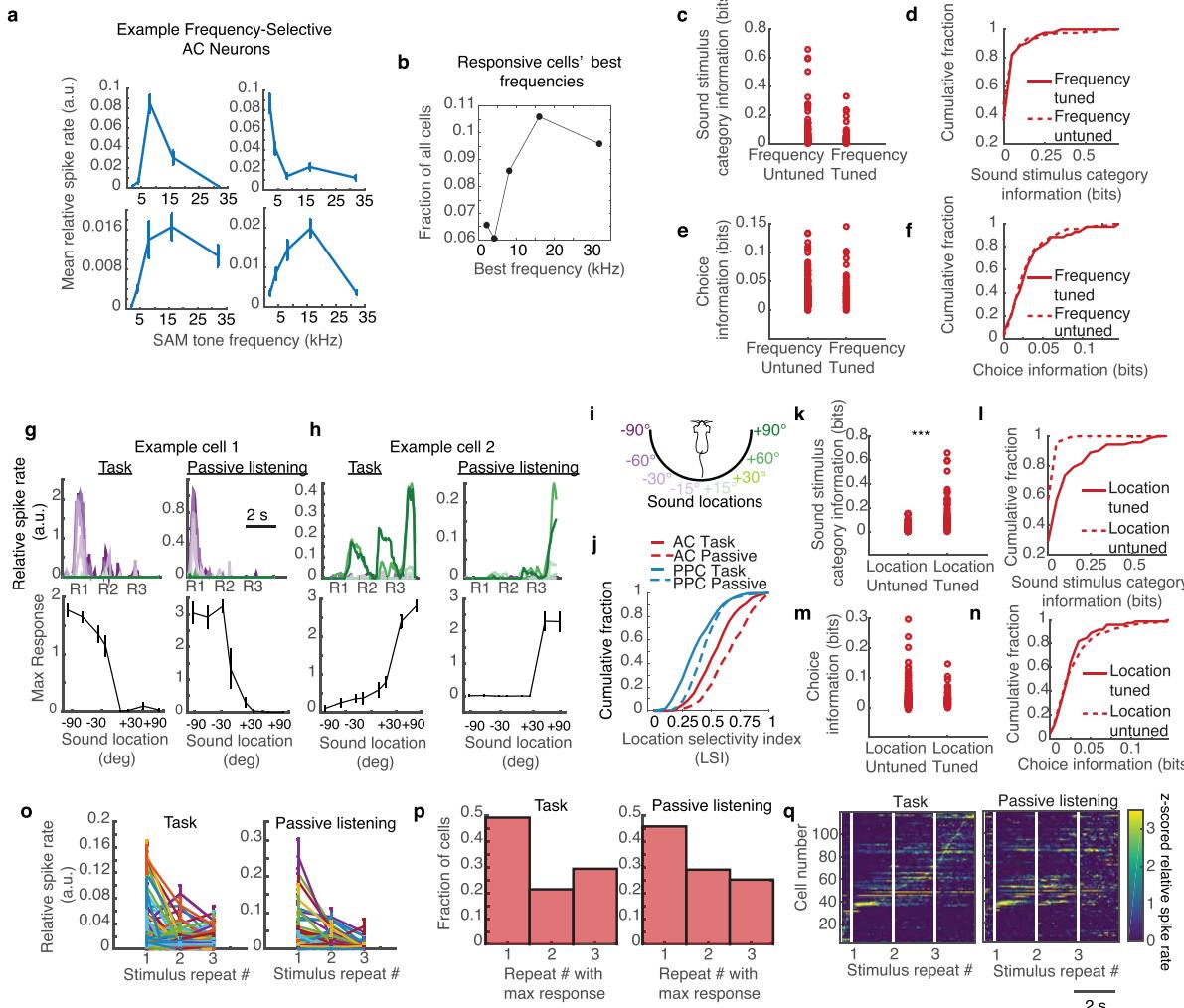
Average posterior correlation was computed for all lags between 0 and 2 s. The dependence of the average posterior correlation on the time lag was fitted with single and double exponential decay functions, defined as $y(t) = \exp(-t/\tau)$ and $y(t) = a \exp(-t/\tau_1) + (1-a)\exp(-t/\tau_2)$, respectively, using a nonlinear least-squares procedure (MATLAB's lsqnonlin function), and the Bayesian information criterion selected the double exponential fit for all conditions considered. Confidence bounds on the fitted parameters were derived from the Jacobian of the exponential function at the best-fit solution via MATLAB's nlpaci function. The value of τ_2 , the component of the fit capturing longer timescale dynamics, is reported in Fig. 4f, h, j. To assess the significance of the differences between two estimated time constants $\tau_2^{(1)}$ and $\tau_2^{(2)}$, the covariance matrix of each estimator $\tau_2^{(i)}$ was computed as $V_i = (\text{RSS}/df)(J_i^T J_i)^{-1}$, where RSS is the residual sum of squares, df is the number of degrees of freedom of the fit (given by the number of lags considered minus the number of parameters of the function, which was 3 for the double exponential function), and J is the Jacobian of the function at the best-fit

solution. The variance σ_i^2 of $\tau_2^{(i)}$ was then determined as the appropriate element on the diagonal of V_i , and a z-test was performed on the difference $\tau_2^{(1)} - \tau_2^{(2)}$ (with the standard deviation $(\sigma_1^2 + \sigma_2^2)^{1/2}$) versus the null value of zero. Holm–Bonferroni correction was applied to control for multiple comparisons.

Single-cell information timescale. For each cell, instantaneous information at lag t from the peak was computed as $(I(t_{\text{peak}} + t) + I(t_{\text{peak}} - t))/2$. Values for all lags between 0 and 0.6 s were averaged across all informative cells (peak information > 0.06 bits), and a noise baseline (0.03 bits) was subtracted from the average information. The resulting average was then maximum-normalized and fitted with single and double exponential decay functions, following the same procedure outlined above for the consistency of the population decoder posterior (Fig. 2i–k). The Bayesian information criterion did not justify using the double exponential fit for AC/choice; hence the single exponential form was chosen across all three conditions considered (AC/stimulus, AC/choice, PPC/choice). Confidence intervals and significance tests for differences between time constants were determined as above for the consistency of the population decoder posterior, with the exception that the number of degrees of freedom for the fits was taken to be 1 rather than 3, following the choice of a single exponential rather than double exponential functional form.

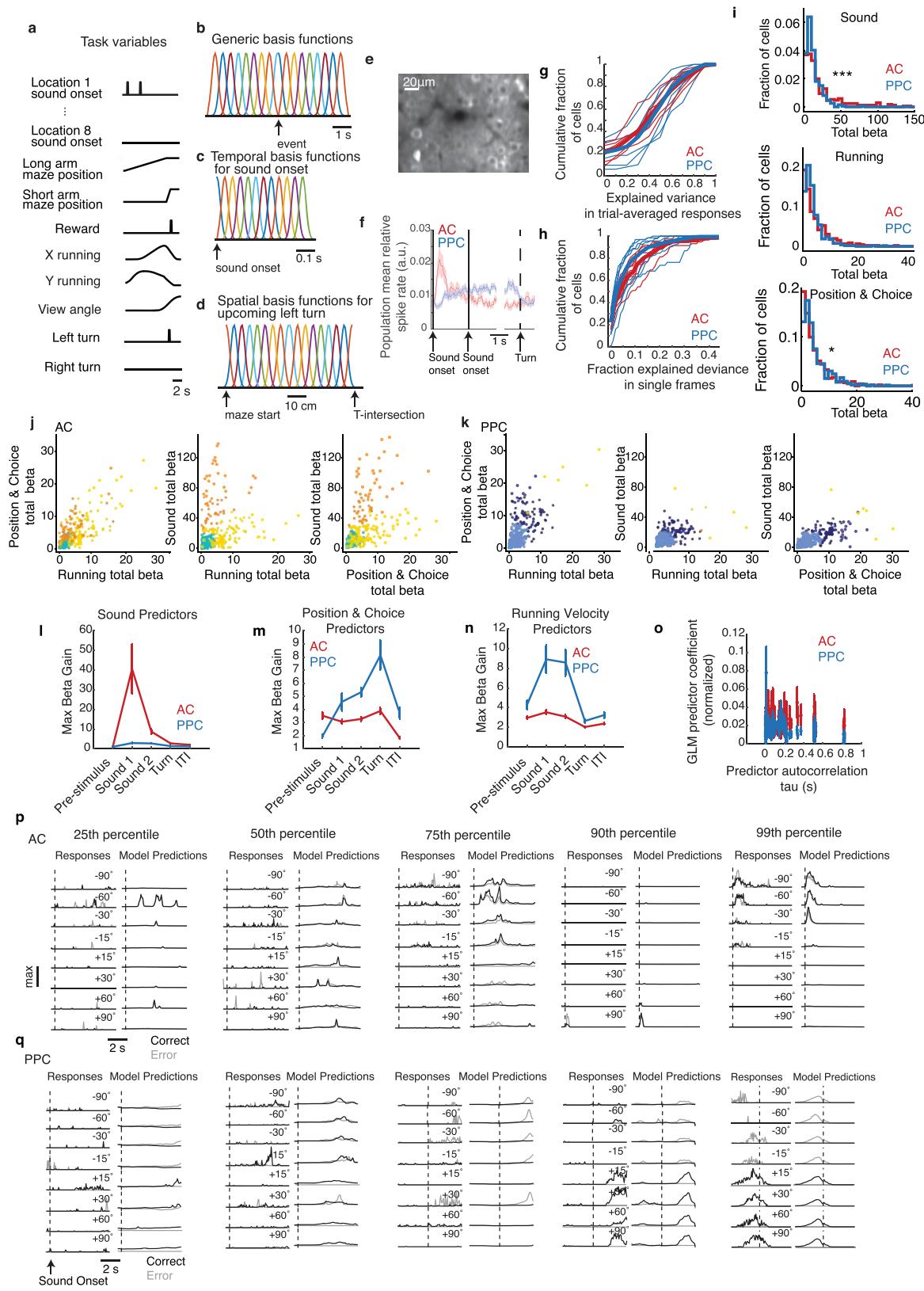
Data and code availability. The data and code that support the findings of this study are available from the corresponding authors upon request.

31. Harvey, C. D., Collman, F., Dombeck, D. A. & Tank, D. W. Intracellular dynamics of hippocampal place cells during virtual navigation. *Nature* **461**, 941–946 (2009).
32. Aronov, D. & Tank, D. W. Engagement of neural circuits underlying 2D spatial navigation in a rodent virtual reality system. *Neuron* **84**, 442–456 (2014).
33. Elhilali, M., Fritz, J. B., Klein, D. J., Simon, J. Z. & Shamma, S. A. Dynamics of precise spike timing in primary auditory cortex. *J. Neurosci.* **24**, 1159–1172 (2004).
34. Pulkki, V. Virtual sound source positioning using vector base amplitude panning. *J. Audio Eng. Soc.* **45**, 10 (1997).
35. Creutzfeldt, O., Hellweg, F. C. & Schreiner, C. Thalamocortical transformation of responses to complex auditory stimuli. *Exp. Brain Res.* **39**, 87–104 (1980).
36. Greenberg, D. S. & Kerr, J. N. Automated correction of fast motion artifacts for two-photon imaging of awake animals. *J. Neurosci. Methods* **176**, 1–15 (2009).
37. Vogelstein, J. T. et al. Fast nonnegative deconvolution for spike train inference from population calcium imaging. *J. Neurophysiol.* **104**, 3691–3704 (2010).
38. McCullagh, P. & Nelder, J. A. *Generalized Linear Models* 2nd edn, Ch. 4 (Chapman and Hall, 1989).
39. Agresti, A. *Categorical Data Analysis* 3rd edn (Wiley, 2013).
40. Lee, D. D. & Seung, H. S. Learning the parts of objects by non-negative matrix factorization. *Nature* **401**, 788–791 (1999).
41. Friedman, J., Hastie, T. & Tibshirani, R. Regularization paths for generalized linear models via coordinate descent. *J. Stat. Softw.* **33**, 1–22 (2010).
42. Quiroga, R. & Panzeri, S. Extracting information from neuronal populations: information theory and decoding approaches. *Nat. Rev. Neurosci.* **10**, 173–185 (2009).
43. Shannon, C. E. A mathematical theory of communication. *ATT Bell Lab. Tech. J.* **27**, 379–423 (1948).
44. Panzeri, S. & Treves, A. Analytical estimates of limited sampling biases in different information measures. *Network* **7**, 87–107 (1996).


Extended Data Figure 1 | Sound frequency tuning in AC neurons.

a, Mean responses (maximum relative spike rate across the 1 s sound presentation) to sinusoid amplitude-modulated pure tones in example AC neurons. These tones were presented to passively listening mice after the task. a.u., arbitrary units. b, Histogram of sound responsive cells' best frequencies, the frequency of the maximum response for each neuron ($n = 206$ neurons, unresponsive neurons were not included). c–f, Information about the sound stimulus category and the mouse's choice in the task were compared between neurons that were untuned or tuned for sound stimulus frequency as measured in a. Significant tuning was defined by comparing the frequency selectivity index ($y_{\text{max}} - y_{\text{mean}})/(y_{\text{max}} + y_{\text{mean}})$, where y_{max} is the mean response to the best frequency, and y_{mean} is the mean response to the other frequencies, with the frequency selectivity index calculated with shuffled trial identities. Frequency-tuned and untuned neurons did not contain significantly different amounts of information about the stimulus category or choice in the task ($P > 0.5$, rank-sum test). g, h, In a subset of imaging experiments ($n = 3$), we played the same sound location stimuli as in the task, in a similar repeating pattern as mice experienced during task trials (three consecutive stimulus repeats). Trial-averaged responses to sound location stimuli measured during the task (left) and during passive listening (right) contexts. Line colours indicate the sound location (see i). Bottom row: tuning curves measured as the average maximal relative spike rate during the sound presentation at each sound location in task (left) and passive (right) contexts. i, Sound location colour legend, which applies to g and h. j, Cumulative distributions of sound location selectivity indices (LSI: $(y_{\text{max}} - y_{\text{mean}})/(y_{\text{max}} + y_{\text{mean}})$, where y_{max} is the mean response to the best location, and y_{mean} is the mean response to the other locations) measured in AC and PPC neurons during the task (solid lines) and passive listening (dashed lines). AC cells had significantly higher LSIs than PPC cells ($P < 0.001$, rank-sum test), and AC cells had significantly higher LSIs in the passive context than the active context ($P < 0.001$, signed-rank test; AC: $n = 329$; PPC: $n = 386$). k, Sound stimulus category information during the task in neurons untuned or tuned for sound location, determined by comparing LSIs in real and shuffled data during passive listening. Neurons with significant sound location tuning had more information about the sound location stimulus category (left versus right), $P < 0.001$, rank-sum test. l, Cumulative distributions of sound category information for neurons tuned and untuned for sound location (using LSI significance). m, Choice information in neurons untuned or tuned for the sound location. n, Cumulative distributions of choice information for neurons tuned and untuned for the sound location (using LSI significance). Location-selective neurons had similar distributions of choice information ($P > 0.5$, rank-sum test). o, Mean response of all neurons across each stimulus repeat during the task (left) and passive (right) contexts, from the subset of imaging experiments with identical passive and active stimuli ($n = 117$ neurons). Responses to sound repeat 1 tended to be higher than responses to repeats 2 and 3 ($P < 0.001$, signed-rank test). p, Histograms of the stimulus repeat during which cells had their maximal responses during task (left) and passive listening (right) contexts. q, Z-scored, trial-averaged activity of all AC neurons with three stimulus repeats in the passive context, sorted by time of peak mean activity and aligned to the time of the first sound onset. Responses during the task (left) and passive listening (right) were sorted by the time of peak response during the task. White vertical lines show the onset times of the first, second, and third sound stimulus repeats. Task trials with more or fewer than three repeats were excluded. The overall temporal pattern of activation across the AC population appeared similar in the two contexts, with a subset of neurons responding during the first sound stimulus presentation, and other neurons responding later, with some responses appearing to depend on subsequent sound stimulus repeats. Many neurons did not appear obviously responsive to the sound stimuli used in the task. Data are, in a, g, h, o, mean \pm s.e.m.

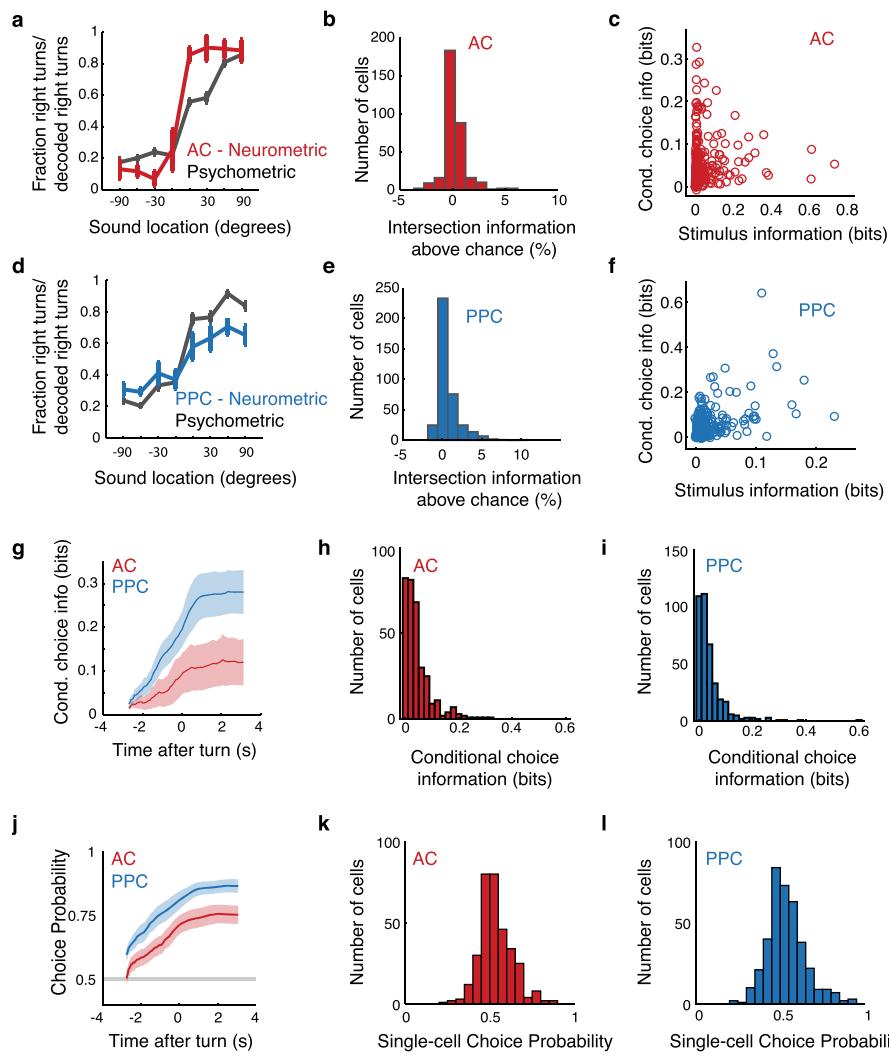
signed-rank test; AC: $n = 329$; PPC: $n = 386$). k, Sound stimulus category information during the task in neurons untuned or tuned for sound location, determined by comparing LSIs in real and shuffled data during passive listening. Neurons with significant sound location tuning had more information about the sound location stimulus category (left versus right), $P < 0.001$, rank-sum test. l, Cumulative distributions of sound category information for neurons tuned and untuned for sound location (using LSI significance). m, Choice information in neurons untuned or tuned for the sound location. n, Cumulative distributions of choice information for neurons tuned and untuned for the sound location (using LSI significance). Location-selective neurons had similar distributions of choice information ($P > 0.5$, rank-sum test). o, Mean response of all neurons across each stimulus repeat during the task (left) and passive (right) contexts, from the subset of imaging experiments with identical passive and active stimuli ($n = 117$ neurons). Responses to sound repeat 1 tended to be higher than responses to repeats 2 and 3 ($P < 0.001$, signed-rank test). p, Histograms of the stimulus repeat during which cells had their maximal responses during task (left) and passive listening (right) contexts. q, Z-scored, trial-averaged activity of all AC neurons with three stimulus repeats in the passive context, sorted by time of peak mean activity and aligned to the time of the first sound onset. Responses during the task (left) and passive listening (right) were sorted by the time of peak response during the task. White vertical lines show the onset times of the first, second, and third sound stimulus repeats. Task trials with more or fewer than three repeats were excluded. The overall temporal pattern of activation across the AC population appeared similar in the two contexts, with a subset of neurons responding during the first sound stimulus presentation, and other neurons responding later, with some responses appearing to depend on subsequent sound stimulus repeats. Many neurons did not appear obviously responsive to the sound stimuli used in the task. Data are, in a, g, h, o, mean \pm s.e.m.



Extended Data Figure 2 | See next page for caption.

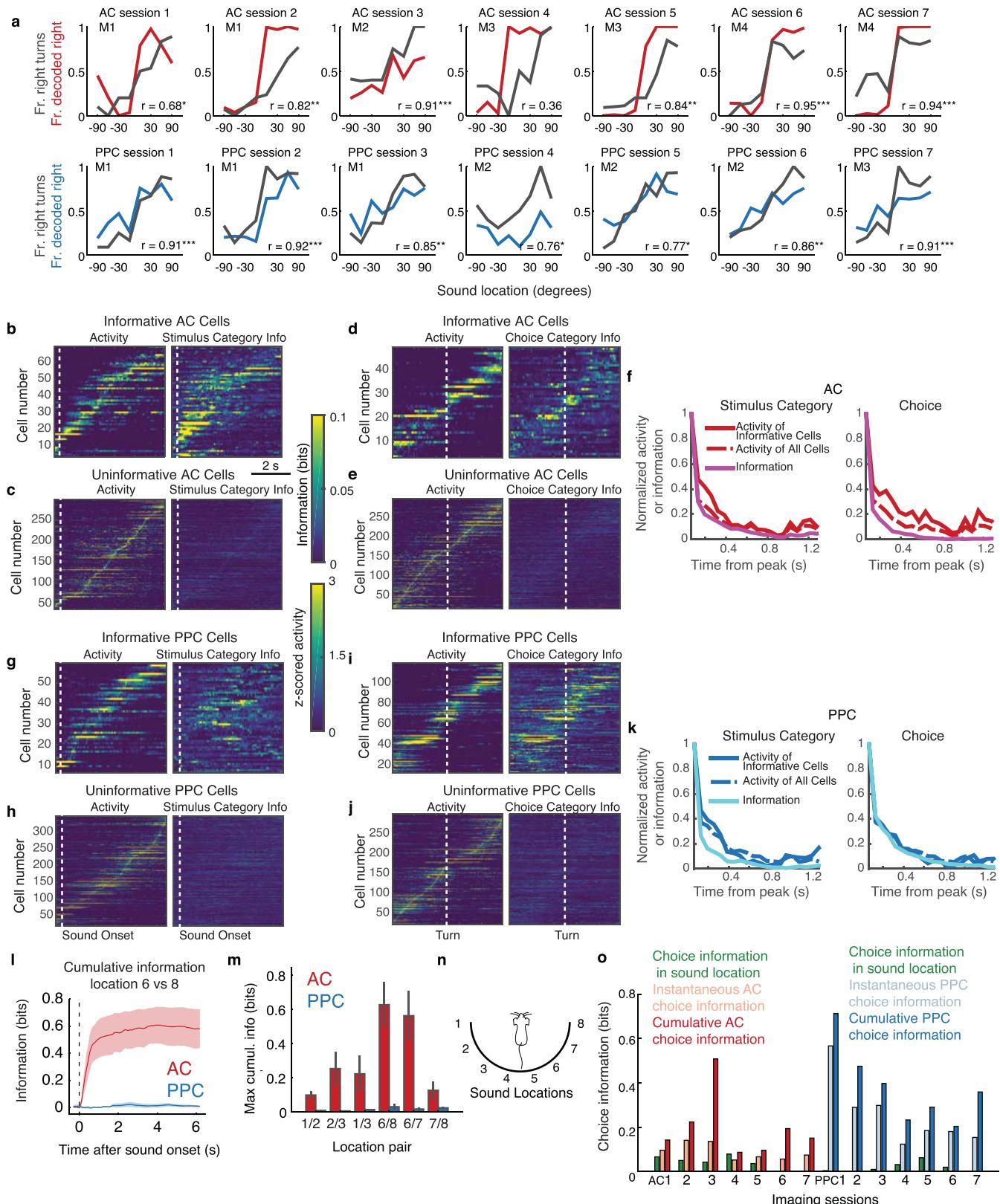
Extended Data Figure 2 | GLM components, fit quality, and model fit examples. **a**, Time course of the behaviour variables included in the GLM during an example trial, in which the sound stimulus was played from location 1 (-90°) and the mouse turned left at the T-intersection to receive a reward. Each predictor was convolved with a set of basis functions (Methods, examples in **b–d**). **b**, Generic set of basis functions that were convolved with behaviour variables and events, extending backwards and forwards in time to model a neuron's response to, and prediction of, events. The density, temporal extent, and width of basis functions were tailored to each behavioural variable (Methods). **c**, Example set of basis functions used for sound onset for each of eight locations. These basis functions are shown before convolving with a vector that specifies sound onset time as a time series of zeros and ones (as shown in **a**). **d**, Example set of basis functions defined spatially along the extent of the stem of the T-maze, which were positive only when the mouse made a left turn in that trial. **e**, Example field of view showing GCaMP6 expression in PPC. **f**, Mean activity across all AC (red) and PPC (blue) cells, across all trials. Shading indicates s.e.m.; $n = 7$ datasets for AC and PPC. **g**, Uncoupled encoding model performance measured using trial-averaged responses and predictions, quantified as explained variance. Each thin line is the distribution from a single dataset ($n = 7$ AC datasets, $n = 7$ PPC datasets). Thick lines indicate mean distributions across datasets. AC versus PPC: $P > 0.1$, rank-sum test. **h**, Uncoupled encoding model performance measured as the fraction of additional explained deviance compared with the null model on frame-by-frame activity (Methods). Each thin line is the distribution from a single dataset. Thick lines indicate mean distributions across datasets. AC versus PPC: $P < 0.05$, rank-sum test. Note that explained deviance is calculated over single imaging frames and single trials in the test dataset (not on averaged data), and, because of trial-to-trial variability of neuron responses, does not approach perfect prediction (1.0). **i**, Histograms of distributions of total beta score (the sum of the absolute value of beta coefficients) in fitted models across all AC and PPC neurons, for predictors in three categories: (1) sound, (2) running, and (3) position/choice (turn direction in the maze). For ease of display, identically zero values were ignored when making the histograms. AC neurons tended to have stronger weights for predictors related

to the sound stimulus ($P < 0.001$, rank-sum test), while PPC neurons tended to have stronger weights for predictors related to position/choice ($P < 0.05$, rank-sum test). **j, k**, Relationship of predictor weights within single neurons in AC and PPC. For each neuron, a score within each predictor category was calculated as the sum of the absolute value of coefficients ('total beta'). Clustering was used to reveal functional groups of neurons tuned to different sets of parameters (see Supplementary Information). Different clusters are indicated with different colours. Note that clustering was performed separately on AC and PPC data, hence the clusters obtained for AC are not related to those for PPC. **l–n**, Fitted model components' gain⁴³ for variables related to running speed, position in the maze and turn direction, and sound location ($\exp(\beta X)$, right, where β is the fitted coefficient, and X is the task predictor convolved with the basis functions), during the various epochs of the trial. Data are mean \pm s.e.m. **o**, Timescales of task predictors. A single exponential was fitted to the autocorrelation of each of the 419 task predictors used in the uncoupled model, to estimate each variable's timescale of variability. Across all AC (red, $n = 329$) and PPC (blue, $n = 386$) neurons, the mean coefficient magnitude fitted by the model for each task predictor was compared with the decay time constant, to determine whether the longer coding timescale in PPC could be explained by neurons in PPC responding preferentially to variables with longer timescales of variability. Even for task variables with longer timescales, AC tended to have greater coefficients ($P < 0.001$, rank-sum test). It was thus unlikely that the longer coding timescales that we measured in PPC were due to its modulation by task variables with long timescales. Data are mean \pm s.e.m. **p**, Example AC responses and model predictions. Trial-averaged responses of example AC neurons during the sound stimulus presentation in correct (black) and error (grey) test trials (left) and the model's predicted responses in data that were left out (right). All responses were normalized to the maximum response across all trial conditions (y axis scale) and aligned to the time of the first sound stimulus onset. Each row is the trial-averaged response or prediction for trials of one of the eight sound location conditions. Neurons with model-fitted performance across the spectrum are included, from poorly fitted (25th percentile), to very well fitted (99th percentile). **q**, Same as **p**, for PPC.



Extended Data Figure 3 | Choice probability and behavioural relevance of sensory information. **a**, Red: AC neurometric curve computed from the performance of a sound location decoder (Supplementary Information). Dark grey: psychometric curve for mice from AC imaging sessions. Session-averaged psychometric and neurometric curves were positively correlated ($r = 0.93$, $P < 0.001$). Data are mean \pm s.e.m. **b**, Intersection information above chance levels in AC neurons. The amount of intersection information per cell above chance level was $0.2\% \pm 0.1\%$ in AC (larger than zero with $P < 0.0001$, paired signed-rank test). **c**, Conditional choice information plotted against stimulus information for all single AC cells. Stimulus information was computed without discounting the correlation between stimulus and choice (Spearman's $r = 0.33$, $P < 0.001$; Supplementary Information). **d-f**, Same as **a-c**, except for PPC. **d**, Session-averaged psychometric and neurometric curves were positively correlated ($r = 0.99$, $P < 0.001$). **e**, Intersection information above chance levels in PPC neurons was

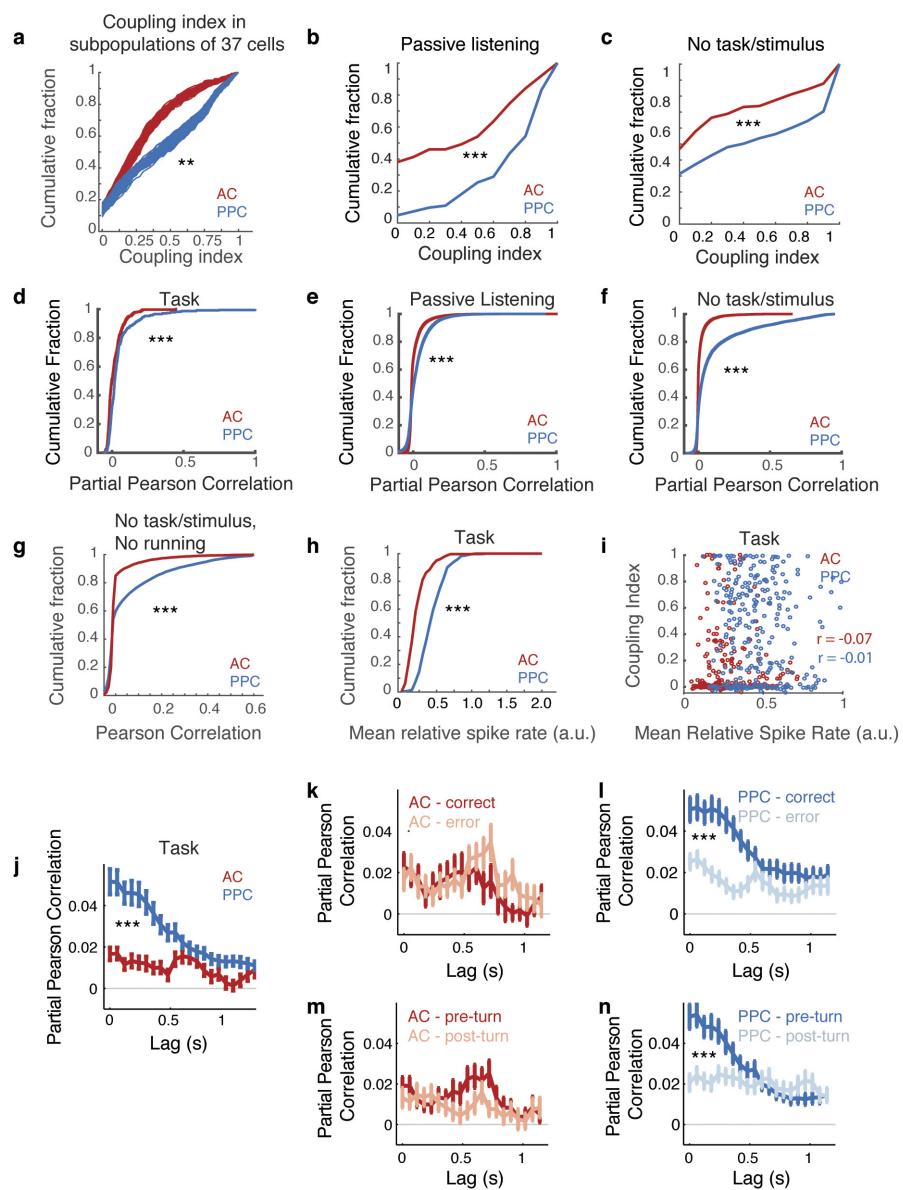
$0.7\% \pm 0.2\%$ ($P < 0.0001$, paired signed-rank test). **f**, Spearman's $r = 0.43$, $P < 0.001$. **g**, Cumulative conditional choice information, computed by the performance of a choice decoder assuming knowledge of all GLM predictors not directly related to choice (Supplementary Information). Cumulative conditional choice information was significant in both AC and PPC (AC: $P < 0.05$, PPC: $P < 0.001$, one-tailed t -test on the value of the choice information at the moment of the turn). Red, AC; blue, PPC; data are mean \pm s.e.m. **h, i**, Total single-cell conditional choice information, computed as the maximum of the cumulative conditional choice information for each cell in AC and PPC. **j-l**, Same as **g-i**, quantifying the performance of the conditional choice decoder as a model-based choice probability¹⁹. **j**, Population choice probability was significantly larger than 0.5 in both AC and PPC at the last aligned time frame ($P < 0.001$, one-tailed t -test). **k**, Mean single-cell choice probability 0.537 ± 0.005 (larger than 0.5, $P < 0.001$, t -test). **l**, Mean single-cell choice probability 0.536 ± 0.006 (larger than 0.5, $P < 0.001$, t -test).



Extended Data Figure 4 | See next page for caption.

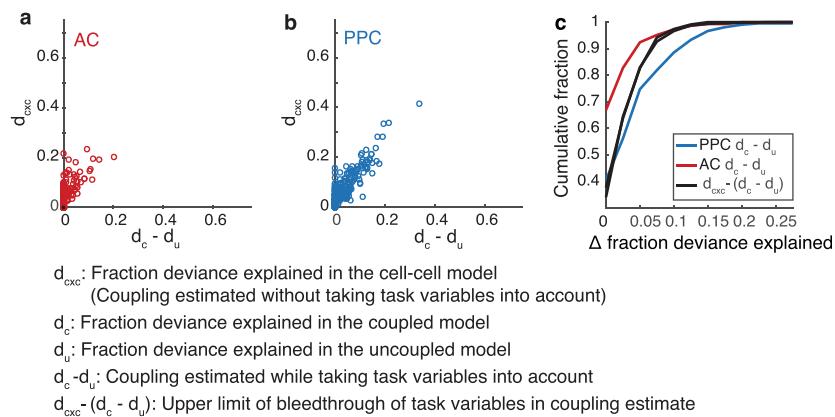
Extended Data Figure 4 | Decoder controls. **a**, Grey: behaviour performance during single imaging sessions used in all analyses. Overall behaviour performance did not correlate with the number of choice-selective neurons in the AC ($r = -0.34, P > 0.1$) or PPC ($r = 0.38, P > 0.1$) populations. Red, blue: single-session neurometric curves using a sound location decoder in AC, PPC, respectively (same as in Extended Data Fig. 3a, d and Supplementary Information). The r value in each panel reports the correlation coefficient between the neurometric and psychometric curves. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$; one-tailed test with a null hypothesis that correlation is not higher than chance. **b**, Left: z -scored trial-averaged activity of all AC neurons with >0.06 bits of stimulus information, sorted by time of peak mean activity, aligned to trial events as in Fig. 1d, e. Right: instantaneous stimulus category information in the same neurons. The scale bar showing time below the top right panel applies to all panels **b–e** and **g–j**. **c**, Left: z -scored, trial-averaged activity of all AC neurons with ≤ 0.06 bits of stimulus information. Right: instantaneous stimulus category information in the same neurons. **d**, Left: z -scored, trial-averaged activity of all AC neurons with >0.06 bits of choice information, sorted by time of peak mean activity. Right: instantaneous choice information for the same cells. **e**, Left: z -scored, trial-averaged activity of all AC neurons with ≤ 0.06 bits of choice information, sorted by time of peak mean activity. Right: instantaneous choice information for the same cells. **f**, Information about stimulus category (left, magenta) and choice (right, magenta) averaged across all AC neurons with ≥ 0.06 bits of information, as a function of the time

from the peak. Normalized, aligned activity in all cells (dashed line) and informative cells (solid red line) are superimposed. **g–k**, Same as **b–f**, for PPC cells. **l–n**, Information about the exact location of the sound stimulus. **l**, The cumulative population decoder was used on subsets of trials from only two sound locations such that there were equal numbers of trials from each location and no other locations were present. The data are shown for a comparison of locations 6 and 8. Note that locations 6 and 8 are part of the same category and indicate the same correct choice in the task. **m**, Maximum cumulative information calculated as in **l** for other location pairs belonging to the same stimulus category (left or right). All imaging experiments were performed in the left hemisphere, so AC had higher information about contralateral sound stimulus locations. Data are mean \pm s.e.m. **n**, Diagram showing the sound location arrangements for location pairs compared in decoders in **l** and **m**. **o**, Choice information in all datasets. Green bars: choice information that could be extracted from sound location owing to an uneven distribution of errors across sound locations. This was a concern, because more errors occurred at sound locations close to the midline (Fig. 1c), and perhaps location tuning could have led to an aberrant choice of information measurement. Light red (AC) and light blue (PPC): maximum instantaneous choice information in neuronal population activity. Dark red (AC) and dark blue (PPC): total cumulative choice information in neuronal population activity. The uneven distribution of error trials across sound locations was not sufficient to explain choice information in AC or PPC.



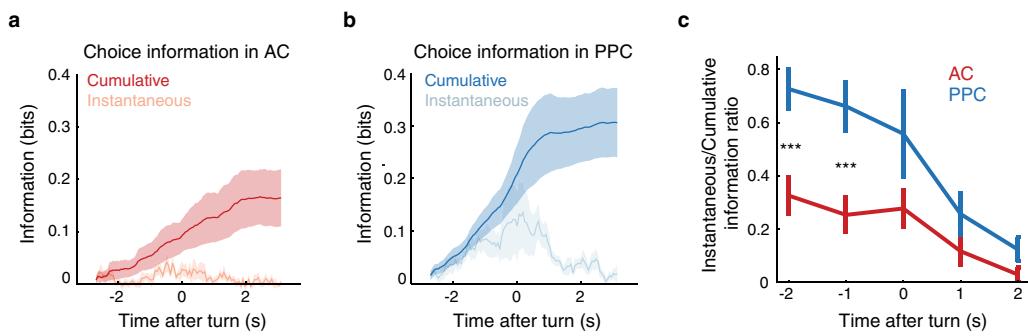
Extended Data Figure 5 | Coupling and Pearson's correlations across contexts. **a**, The coupled model was fitted using random subpopulations of 37 neurons (the number of neurons in the smallest dataset) 100 times in all AC and PPC datasets. Cumulative distributions of performance of the coupled model using all subsamples of AC (red) and PPC (blue) datasets show that all PPC datasets still had greater coupling than all AC datasets ($P < 0.01$, rank-sum test), even when using smaller numbers of neurons. **b, c**, Cumulative distributions of the coupling index in AC (red) and PPC (blue), measured during passive listening and the ‘no task/stimulus’ context. **d**, Cumulative distributions of partial Pearson’s correlations (Methods) during the pre-turn period of the task in AC (red) and PPC (blue). Consistent with the higher coupling indices measured in PPC (Fig. 3d), correlations were higher in PPC than AC ($P < 0.001$, Kolmogorov–Smirnov test). **e**, Cumulative distributions of partial Pearson’s correlations during passive listening to the stimulus sets used in the task in AC (red) and PPC (blue) datasets. PPC was more correlated than AC, even when the mouse was not engaged in a task ($P < 0.001$; Kolmogorov–Smirnov test). **f**, Cumulative distributions of partial Pearson’s correlations measured in the ‘no task/stimulus’ context in AC (red) and PPC (blue). Again, PPC was more correlated than AC, in the absence of task or sound stimulus presentations ($P < 0.001$, Kolmogorov–Smirnov test). **g**, Cumulative distributions of Pearson’s correlations measured in the

‘no task/stimulus’ context when the mice were stationary (not running) on the ball (AC versus PPC: $P < 0.001$, Kolmogorov–Smirnov test). **h**, Cumulative distributions of mean relative spike rates in AC (0.24 ± 0.14 arbitrary units, mean \pm s.e.m.) and PPC (0.45 ± 0.18 arbitrary units). AC versus PPC: $P < 0.001$, Kolmogorov–Smirnov test. **i**, Coupling index versus mean relative firing rate measured during the pre-turn period of the task. Each dot is one neuron. Because coupling and firing rate were not positively correlated (AC: $r = -0.020$; PPC: $r = 0.007$; $P > 0.25$), it is unlikely that higher firing rates in PPC caused an artefactual increase in correlation and coupling relative to AC. **j**, Partial Pearson’s correlations, computed across trials for time lags spanning 0 to over 1 s in AC (red) and PPC (blue), over all data, during the task. *** $P < 0.001$, rank-sum test on the average partial Pearson’s correlation across lags smaller than 0.5 s. Data are mean \pm s.e.m. across cell pairs. **k**, Partial Pearson’s correlations in AC populations, computed separately for correct (red) and error (pink) trials, for pre-turn data. Data are mean \pm s.e.m. across cell pairs. **l**, Same as **k**, for PPC (dark blue, correct trials; light blue, error trials). *** $P < 0.001$, rank-sum test on the average partial Pearson’s correlation across lags smaller than 0.5 s. **m**, Partial Pearson’s correlations in the pre-turn (red) and post-turn (pink) trial epochs in AC populations. Data are mean \pm s.e.m. across cell pairs. **n**, Same as **m**, for PPC. *** $P < 0.001$, rank-sum test on the average partial Pearson’s correlation across lags smaller than 0.5 s.



Extended Data Figure 6 | Using the cell-cell model to test possible contributions of task features to coupling. To help exclude the possibility that coupling parameters simply allowed the model to explain behavioural correlates not included in the uncoupled model, rather than neuron–neuron correlations, we removed all the task predictors from our model to create a cell–cell model that only had coupling predictors. We reasoned that if the model could misattribute common drive to neurons by behaviour variables, then the cell–cell model should be able to take on the uncoupled model’s prediction of responses to task parameters. We estimated an upper bound on the bleed-through of task variables to coupling parameters by comparing the cell–cell model d_{cxc} with $d_c - d_u$, the increase in model performance when including coupling predictors in addition to the task predictors. If the coupling predictors could explain all of the responses related to the task predictors, d_{cxc} would far exceed the

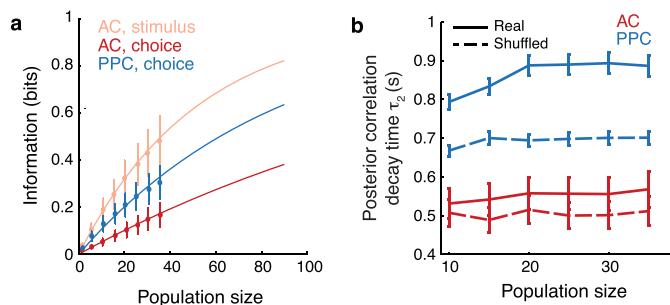
coupling value. **a, b**, Performance of a version of the encoding model with only the activity of other neurons as predictors and no task predictors (cell–cell model) compared with coupling (performance of coupled model – performance of uncoupled model) in AC (**a**) and PPC (**b**). **c**, Comparing cumulative distributions of coupling in AC (red line) and PPC (blue line) with our estimates of an upper bound on the extent to which coupling could be explained entirely by task-related variables (black lines). Note that the coupling distribution in AC cells (red line) was mostly restricted to values less than the upper bound on coupling explainable by task-related variables, while coupling in many PPC neurons (blue line) exceeded it ($P < 0.01$, rank-sum test). These results suggest that the higher level of coupling measured in PPC is unlikely to be due to shared common inputs to PPC neurons relating to task variables.



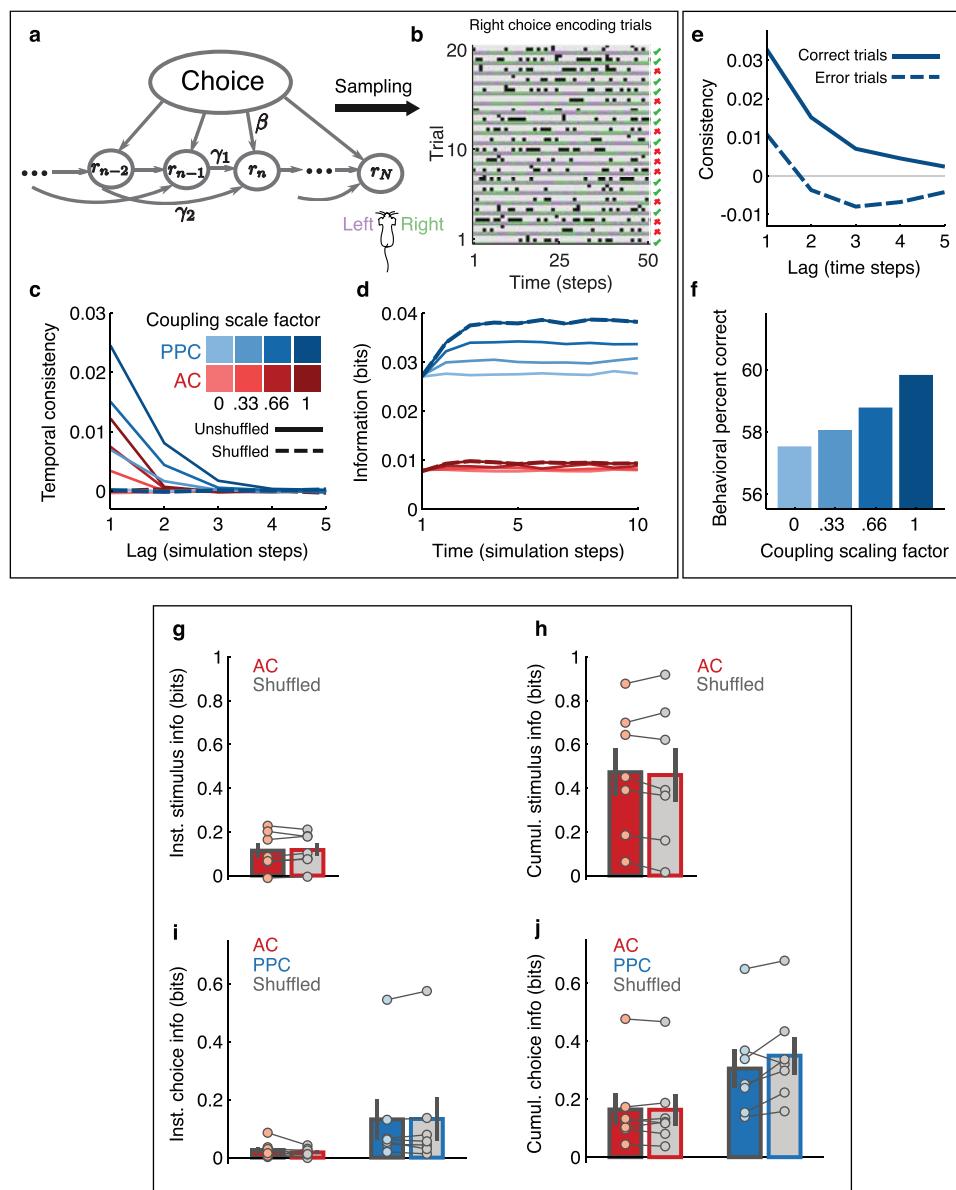
Extended Data Figure 7 | Choice information redundancy.

a, b, Cumulative (dark lines) and instantaneous (light lines) choice information in AC ($n=7$ datasets) and PPC ($n=7$ datasets), aligned to the turn and averaged across datasets. Data are mean \pm s.e.m. across datasets. **c**, Ratio of instantaneous to cumulative choice information in 1 s windows, relative to the time of the mouse's turn in the maze. In PPC, instantaneous

and cumulative information were similar before the turn. These results support our findings in Fig. 4, that information in PPC was consistent (in other words, redundant) across time before the mouse reported its choice by turning in the maze. Data are mean \pm s.e.m across datasets. *** $P < 0.001$, z -test.



Extended Data Figure 8 | Population scaling of information and consistency. **a**, Scaling of total information content (maximum of cumulative information) with population size. Data are mean \pm s.e.m. across datasets. Lines: analytical prediction from the random overlap model (Supplementary Information). **b**, Scaling of information consistency with population size, by measuring τ_2 (the long timescale component of decoder consistency, as in Fig. 4f) while varying population sizes (Supplementary Information). Data are mean \pm s.e.m. across datasets. While information in AC and PPC grew with increasing population size, the coding timescale remained constant in AC, but grew modestly in PPC.



Extended Data Figure 9 | Simple generative model of sequential neuronal activity and the effect of shuffling on information content.

a, Schematic of simple model of choice tuning and statistical coupling between cells, expressed as a probabilistic graphical model. 'Choice' indicates the choice encoded by the neuronal population in any given trial, r_i represents the activation of cell i , β is the strength of choice tuning, and γ_1 and γ_2 represent the strength of the statistical coupling of cell i to cell $i - 1$ and cell $i - 2$, respectively (Supplementary Information).

b, Example of cell activity and decoded choice signal generated by the model for a batch of 20 trials encoding right choice. For each trial, the top row indicates the activity of left-preferring cells (black, active). The middle row indicates the activity of right-preferring cells. The bottom row indicates the choice decoded at each instant from the population activity (green, right; purple, left; Supplementary Information). Symbols on the right indicate whether the readout model implemented the correct choice.

c, Temporal consistency computed across 10^5 trials, each composed of 50 time steps. Different shades of red (or blue) indicate different multiplicative scaling factors applied to both coupling parameters γ_i , ranging from 0 for no coupling to 1 for the values derived from experimental data in AC (or PPC). Dashed lines: consistency computed for shuffled data (superimposed with the solid lines corresponding to no coupling).

d, Same as c, for the instantaneous choice information

contained in the choice signal generated by the model, over the first ten time steps of the simulation. Note how cross-cell coupling enables accumulation of choice information. Note also how information in shuffled data is identical to the information in the unshuffled data.

e, Temporal consistency for simulated PPC data, computed separately for behaviourally correct trials (solid line) and error trials (dashed line), as determined by the readout model described in Supplementary Information.

f, Behavioural performance generated by the readout model (Supplementary Information) as a function of strength of the coupling in the model; 0 corresponds to no coupling, 1 to the value of the coupling parameters derived from the experimental data.

g, Instantaneous stimulus information measured 1 s after the first stimulus onset, in real experimental data (red) and after disrupting coupling by shuffling the identities of trials of the same condition independently for each neuron (grey). Information was computed for random subpopulations of 37 cells and averaged across 48 such random selections. Circles represent individual datasets, bars represent the average across datasets, data represent mean \pm s.e.m.

h, Same as g, for cumulative information at the last aligned time point in a trial.

i, Same as g, for choice information in both AC (red) and PPC (blue) at the moment of the turn.

j, Same as i, for cumulative choice information at the last aligned time point in the trial.